THE UNIVERSITY *of* EDINBURGH

The evolution of sex and recombination in large, finite populations

Matthew Hartfield

Submitted for the degree of Doctor of Philosophy

University of Edinburgh

2012

To Fred, Joan, and Chris

– intelligence is hereditary, isn't it?

"So then, what do you believe in?"

"Sex and death - two things that come once in a lifetime...but at least after death, you're not nauseous."

– Woody Allen, *Sleeper*

Acknowledgments

First and foremost I'd like to offer huge thanks to my supervisor Peter Keightley, who's patience, support and persistence over the years has caused my intellectual development to rise in leaps and bounds. With his help I have achieved numerous successes that would not be possible without his motivation, and he was always happy to give me the freedom to pursue personal projects. Massive thanks also to Sally Otto, who co-supervised the simulation work, and accommodated me at the University of British Columbia over the summers of 2009 and 2010. It is with her careful assistance that I was able to understand the complex mathematical work surrounding the evolution of sex and recombination. A third thanks goes to Nick Colegrave, my second supervisor in Edinburgh, who was always on hand to provide careful critiques of my work, and to answer my numerous (and often silly) questions about research in a friendly manner.

Thanks also to the Keightley and Otto groups over the years (Dan Halligan, Lél Eöry, Thanasis Kousathanas, Andrew Morgan, Rob Ness, Fiona Oliver, Adrian Schneider in Edinburgh; Crispin Jordan, Rich FitzJohn, Aleeza Gerstein, Kay Hodgins, Leithen M'Gonigle in Vancouver), as well as the attendees of the Charlesworth lab meetings, for friendly support, advice and assistance. On top of that, thanks go to those who reviewed, assisted with, and gave feedback on various parts of this thesis: Nick Barton, Brian Charlesworth, Denis Roze, Joachim Hermisson, Mike Whitlock, Kayla King, Tom Little, Bill Hill, and various anonymous reviewers.

I would like to thank the Biotechnology and Biological Science Research Council for funding my PhD. I would also like to acknowledge support from an University of British Columbia Graduate Student International Research Mobility Award, which helped me to visit Sally Otto's lab; and the Genetics Society UK and the University of Edinburgh's James Rennie Bequest for funding to assist with conference visits. The

work in Chapters 3 and 5 has made use of the resources provided by the Edinburgh Compute and Data Facility (ECDF) (http://www.ecdf.ed.ac.uk/), which is partially supported by the eDIKT initiative (http://www.edikt.org.uk). Work in Chapter 5 has also made use of computing resources provided by WestGrid and Compute/Calcul Canada.

Thanks to my close family - Fred, Joan, and Chris Hartfield - for continuous moral support and copious amounts of free food during the more stressful periods of my PhD.

Finally, a huge thanks to the various staff and students of the Ashworth Labs, as well as friends outside the department, for making my time in Edinburgh an extremely happy one. It has been an honour to work in such a joyful and intellectually stimulating department; my studies would not have been anywhere near as inspirational otherwise.

Publications

The following papers have arisen from this thesis. Those that have been accepted for publication at time of the completion of the thesis are included in the appendix.

- Hartfield, M., Otto, S.P. and Keightley, P.D., 2010. The Role of Advantageous Mutations in Enhancing the Evolution of a Recombination Modifier. *Genetics*, **184**(4): 1153–1164.
- Hartfield, M. and Otto, S.P., 2011. Recombination and Hitchhiking of Deleterious Alleles. *Evolution*, **65**(9): 2421–2434.
- Hartfield, M. and Keightley, P.D., 2011. Current hypotheses for the evolution of sex and recombination. *Integrative Zoology, accepted (in press)*.

Contents

Thesis abstract

This thesis investigates how breaking apart selection interference ('Hill-Robertson' effects) that arises between linked loci can select for higher levels of recombination. Specifically, it mainly studies how the presence of both advantageous and deleterious mutation affects selection for recombination. These evolutionary advantages are subsequently investigated with regards to sex resisting asexual invasion in a subdivided population.

i) KEIGHTLEY and OTTO (2006) showed a strong advantage to recombination in breaking apart selection interference, if it acts across multiple, linked loci subject to recurrent deleterious mutation. Their model is modified to consider selection acting on recombination if a small proportion of mutations are advantageous. This leads to a greater increase in selection acting on a recombination modifier, compared to cases where only deleterious mutations are present.

ii) Branching-process methods are developed to quantify how likely it is that a deleterious mutant hitchhikes with a selective sweep, and how recombination between the two loci affects this process. This is compared to the neutral hitchhiking model, to determine how levels of linked neutral diversity would differ between the two scenarios. A simple application with regards to human genetic data is provided.

iii) Population subdivision can maintain costly sex, as a consequence of restricted gene flow slowing the spread of invading asexuals, which leads to an excessive accumulation of deleterious alleles. However, previous work did not quantify whether costly sex can be maintained with realistic levels of population subdivision. Simulations in this thesis show that the level of population subdivision (as measured by F_{st}) needed to maintain costly sex decreases with larger population size; however critical F_{st} values found are generally high, compared to surveys of geographically-close populations. The lowest levels of population subdivision that maintained sex were found if mutation is both advantageous and deleterious, and demes were arranged in a one-dimensional stepping-stone formation.

iv) An analytical method is developed to calculate how long it takes an advantageous mutation (such as an invading asexual) to spread through a subdivided population. The flexibility of the methods created means that they can be applied to different types of stepping-stone populations. It is shown how to formulate the fixation time for one-dimensional and two-dimensional structures, with analytical methods showing a good fit to simulation data.

Chapter 1

Current hypotheses for the evolution of sex and recombination

Published as: Hartfield, M. and Keightley, P.D., 2011. Current hypotheses for the evolution of sex and recombination. *Integrative Zoology, accepted (in press)*.

Chapter abstract

The evolution of sex is one of the most important and controversial problems in evolutionary biology. Although sex is almost universal in higher animals and plants, its inherent costs have made its maintenance difficult to explain. The most famous of these is the two-fold cost of males, which can greatly reduce the fecundity of a sexual population, compared to a population of asexual females. Over the last century, multiple hypotheses, along with experimental evidence to support these, have been put forward to explain widespread costly sex. In this review, I outline some of the most prominent theories, along with the experimental and observational evidence supporting these. Historically, there have been four classes of theories: the ability of sex to fix multiple novel advantageous mutants (Fisher-Muller hypothesis); sex as a mechanism to stop the build-up of deleterious mutations in finite populations (Muller's Ratchet); recombination creating novel genotypes that can resist infection by parasites (the 'Red Queen' hypothesis); and the ability of sex to purge bad genomes if deleterious mutations act synergistically ('Mutational Deterministic' hypothesis). Current theoretical and experimental evidence seems to favour the hypothesis that sex breaks down selection interference between new mutants, or it acts as a mechanism to shuffle genotypes in order to repel parasitic invasion. However, there is still a need to collect more data from natural populations and experimental studies, which can be used to test different hypotheses.

1.1 Introduction

What is sex? Sexual reproduction is usually defined as a means of propagation that requires two parents to combine genetic material, usually by uniting two cells (gametes) containing chromosomes from the parents, in order to form a zygote. Before gametes are produced, the parents' genomes first undergo recombination and genetic segregation during meiosis (KLECKNER 1996). Researchers into the evolution of sex are therefore also interested in determining what conditions favour the evolution of recombination, as it is seen as a precursor to the appearance of obligate sex.

This method of reproduction contrasts with asexuality, where in general a parent clones its genotype to reproduce (although there are examples of asexuals undergoing recombination within their own genome (STENBERG and SAURA 2009)). Asexuality is very rare in nature, since only around 0.1% of animal species are obligate asexuals (VRIJENHOEK 1998). Most asexual lineages have recently evolved from sexual predecessors (VRIJENHOEK 1998; SIMON *et al.* 2003), although there may exist a few 'ancient' asexuals, the best-known candidate being the bdelloid rotifers (VRIJENHOEK 1998; MARK WELCH and MESELSON 2000; MARK WELCH *et al.* 2008).

The prevalence of sexual reproduction indicates that there should be a clear and obvious reason as to why it is advantageous. However, this is far from the case: the origin and maintenance of sexual reproduction has remained one of the most elusive questions in evolutionary biology. The reason for this is that sex incurs major costs in comparison to asexual reproduction (MAYNARD SMITH 1978), and to this day no universally accepted explanation exists as to how sex evolved and is maintained in the face of these disadvantages. This review will describe some of the major costs associated with sex, and the most prominent hypotheses that have been put forward to explain its evolution and maintenance.

Sex is a costly endeavour. The most famous of the major costs has been labelled

as the 'twofold' cost of sex. This manifests itself through two outcomes, due to the fact that sexual females invest resources into the production of males, or male gametes in the case of hermaphrodites, which in themselves do not themselves provide any resources to the next generation (see recent review by LEHTONEN *et al.* (2012) for more information). The first and probably the most common usage refers to a 'cost of males' (MAYNARD SMITH 1978), illustrated in Figure 1.1. With biparental sexual reproduction, a male and a female have to meet in order to reproduce. If this results in the birth of one son and one daughter, on average, then the population will be maintained at a constant size. However, in a population of asexuals, energy is invested only in the female function, rather than in both female and male functions. Therefore, parthenogenetic females just need to clone themselves, and males become irrelevant. As a result of this, if each parent produces two female offspring, an asexual population can quickly double in size and easily displace existing sexuals. The twofold cost also refers to a "cost of meiosis" in anisogamous organisms (WILLIAMS 1975; LIVELY and LLOYD 1990), where each sexual parent only contributes half its genes to its offspring, decreasing its genetic contribution and thus the relatedness between parent and offspring.

There are additional costs that can affect the possible emergence of sexual reproduction. Recombination can destroy positive associations between selected clusters of alleles, reducing an individual's fitness, so selection will act against maintaining recombination (NEI 1967). Such a 'recombination load' has been observed in *Drosophila melanogaster* (CHARLESWORTH and CHARLESWORTH 1975). Sexuals also have to expend energy to find mates, and there is the risk of sexual reproduction spreading diseases between parents, or to their offspring (LOCKHART *et al.* 1996; OTTO 2009). However, it has been argued that the cost of males can be decreased through sexual selection (SILLER 2001; AGRAWAL 2001), increased intraspecific competition amongst asexuals (DONCASTER *et al.* 2000), or sexuals increasing their variance in fecundity (BLACHFORD and DOEBELI 2009). Costs to sex have been observed in field studies of

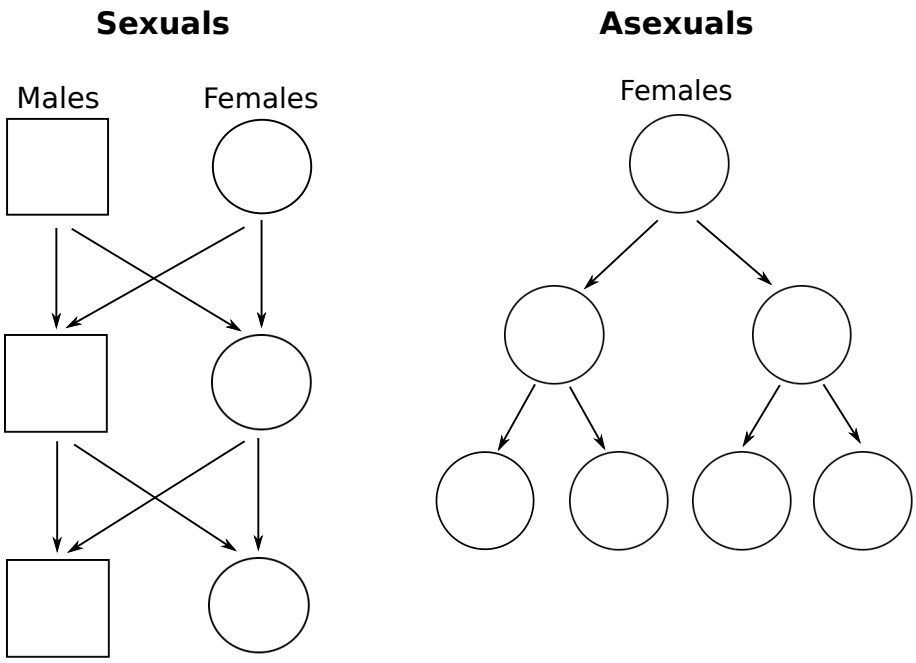


Figure 1.1: A schematic illustrating the twofold cost of males. Males are represented by squares and females by circles.

Antennaria parlinii (MICHAELS and BAZZAZ 1986), *Potamopyrgus antipodarum* snails (JOKELA *et al.* 1997), and psychid moths (KUMPULAINEN *et al.* 2004).

Direct advantages to sex and recombination. Several physiological explanations have been offered to suggest why sex may be advantageous. These 'direct' hypotheses account for the evolution of sex and recombination due to an immediate effect they confer on a individual's fitness, in which these mechanisms act. This is in contrast to 'indirect' hypotheses, which explain the evolutionary advantages of sex and recombination through mixing genetic material from two parents OTTO and LENORMAND (2002).

One direct hypothesis is that sexual reproduction repairs damaged DNA and thus 'regenerates' the genome (BERNSTEIN *et al.* 1988). However, there is little evidence that recombination is essential for DNA repair, even though it could offer an inexpensive way of doing so (MAYNARD SMITH 1988a). Experiments with *Bacillus subtilis* and *Haemophilus influenzae* also failed to find evidence for transformation (and thus recombination) evolving in order to repair damaged DNA (REDFIELD 1993). Subsequent experiments also show that competence and transformation protect *Streptococcus pneumoniae* against non-DNA-damaging processes, indicating that transformation may not necessarily have evolved solely as a mechanism to repair damaged DNA (ENGELMOER and ROZEN 2011). The repair hypothesis also has difficulty explaining the maintenance of sexual reproduction, since double-strand breaks are induced during meiosis in sexuals (KLECKNER 1996). It also does not consider the evolution of asexual diploids, which can utilise the second copy of a specific gene as a template for DNA repair (OTTO and LENORMAND 2002).

Another proximate explanation for sex is that it can improve the transmission of 'selfish' genes (GODDARD *et al.* 2001). In sexual individuals, transposable elements can be passed on to every offspring produced, even if present as a heterozygote, hence their rate of spread will be faster in comparison to non-selfish genes. Transposons that caused sex to evolve could then be selected for as a by-product of ensuring their own

rapid transmission, even if they cause a substantial fitness reduction to the host (HICKEY 1982). However, whilst this theory could explain the initial emergence of sex, it cannot easily explain how sex is maintained. Once selfish genes invade a population and reach a high frequency, asexual individuals can propagate selfish elements just as quickly as sexuals (OTTO and LENORMAND 2002).

Population genetics advantages to sex and recombination. Because such direct, short-term hypotheses have limited power in explaining the evolution of sexual reproduction, much more attention has focused on indirect population genetics based hypotheses instead. The concept underpinning these explanations is that by combining genomes from different backgrounds, sex and recombination create better genotypes that would not be formed asexually. Such fitter sexual individuals are therefore more likely to reproduce and persist in the long term. This idea was memorably summarised by WILLIAMS (1975), who compared the different mating systems acting in a fluctuating environment to buying lottery tickets. Asexuality was akin to buying lots of tickets that all had the same number. Sex, however, was similar to buying greatly fewer tickets, but with each one having different numbers, so it is more likely to produce a 'winner'. The main population-genetics hypotheses can be placed into one of three categories:

1. *Breaking apart interference between selected loci.* Also known as Hill-Robertson interference (HILL and ROBERTSON 1966), this is where selection acting on one locus interferes with selection acting at a second, linked locus in a finite population. Associations between combinations of favoured and disfavoured alleles (creating negative linkage disequilibrium) are formed from the combination of selection and sampling error in finite populations. Recombination breaks apart such interference and improves the response to selection. The classic theories related to this broad idea are the 'Fisher-Muller' hypothesis (FISHER 1930; MULLER 1932), where sexuals can combine beneficial alleles into the same genome, and

'Muller's Ratchet' (MULLER 1964) caused by the irreversible build-up of deleterious mutations in finite asexual populations.

2. *Parasitic resistance ('Red Queen' hypothesis).* By recombining genomes, sexuals are more likely to create new genotypes that are able to adapt to environments that fluctuate deterministically. Negative linkage disequilibrium arises in asexuals, as the fittest genotypes become disproportionally infrequent due to fluctuating epistasis MAYNARD SMITH (1971); CHARLESWORTH (1976). The best-known application of this hypothesis concerns sex as a means to resist parasitic infection (JAENIKE 1978; HAMILTON *et al.* 1990).

3. *'Mutational Deterministic'*. This hypothesis is based on a deterministic model of an infinite population. If the deleterious mutation rate is high enough and deleterious mutants act synergistically (that is, a collection of deleterious mutants cause a greater reduction in log fitness than expected if acting independently), recombination can restore fitness variance that would otherwise decrease due to deleterious mutation accumulation (KONDRASHOV 1982, 1993).

This review will give an overview of each of these theories and their ability to explain the prevalence of sex. Whilst there exist realistic scenarios under which each hypothesis could explain the evolution of costly sex, none have managed to provide a sufficient explanation as to why costly sex is ubiquitous in nature, especially since there has been little empirical testing of some of these hypotheses. This is because it has been hard (until recently) to obtain accurate data against which to test these theories. The difficulty of explaining sex has led to the suggestion that a 'pluralist' approach might provide the best answer, where several processes work in tandem to overcome the two-fold cost of sex (WEST *et al.* 1999). Although WEST *et al.* (1999) considered 'Red Queen' and 'Mutational Deterministic' processes working together, subsequent research has also considered how other mechanisms interact to potentially maintain sex

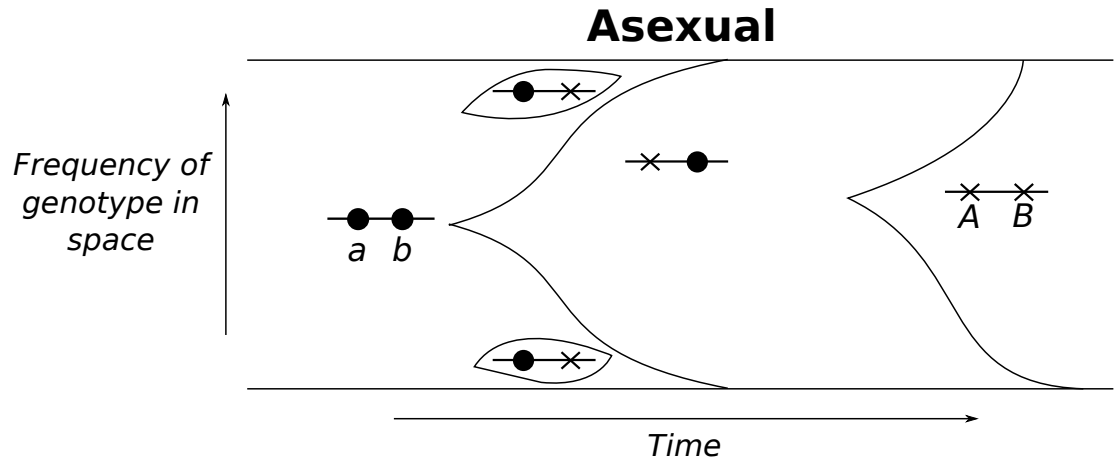
(for examples, see HOWARD and LIVELY (2002) and Chapter 3). Here, I shall separate the discussion of the different hypotheses, due to different mechanisms affecting the evolution of sex in each case (such as finite population size, fluctuating selection or the effects of epistasis).

1.2 Breaking apart selection interference

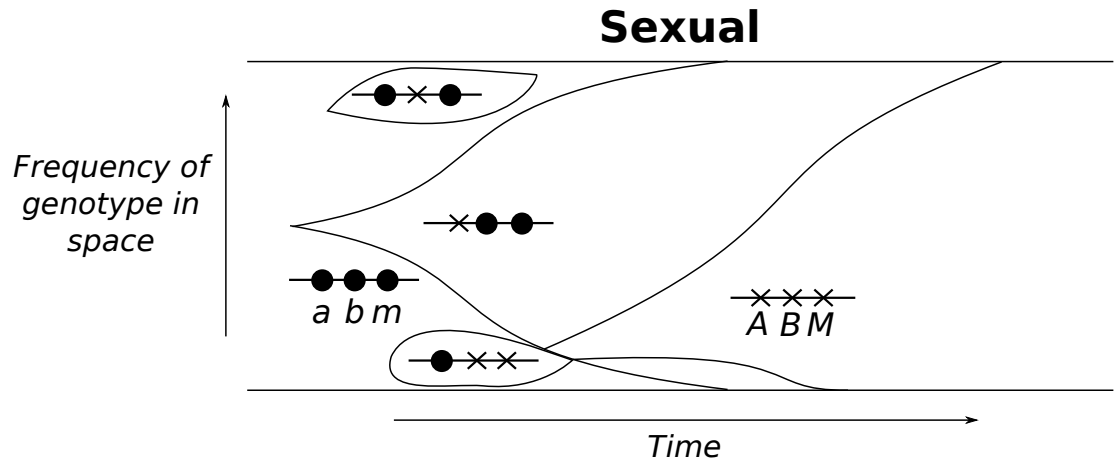
Early theories: 'Fisher-Muller' hypothesis and 'Muller's Ratchet'. In finite asexual populations, sampling and drift can impede the efficacy of selection acting at linked loci. By recombining genomes, the response to selection at individual loci is increased, leading to a higher mean fitness. It could be argued (BURT 2000) that a similar idea was first put forwards by WEISMANN (1887), who contended that sexual reproduction "may be regarded as a source of individual variability, furnishing material for the operation of natural selection". This increase in genome-wide variability thereby causes sex to bring together favourable alleles, whilst less-fit genomes will be 'weeded out'.

FISHER (1930) and MULLER (1932) formalised this argument as applied to the case of two beneficial alleles arising at separate, linked loci. Both argued that sexuals could recombine the two alleles into a new genome, quickly creating a fitter individual. Asexuals, on the other hand, would have to wait for both mutants to arise and fix in sequence in the same lineage, which would take a greatly increased amount of time. This claim was verified by CHRISTIANSEN *et al.* (1998), who determined that in asexual populations, the production of a genome containing both mutants would take on the order of $1/\sqrt{N\mu^2}$ generations, but only $1/\sqrt[3]{N\mu^2}$ in sexual populations, where μ is the mutation rate at each locus and N is the haploid population size. Figure 1.2 outlines a schematic of this process.

Another classical hypothesis to explain the evolution of sex and recombination in finite populations was put forward by MULLER (1964). In nature, he noted that muta-



(a)



(b)

Figure 1.2: How the Fisher-Muller hypothesis works. Figure (a) shows how an asxual lineage will sequentially fix advantageous alleles A and B from initial alleles a and b. Figure (b) shows how a modifier allele for increased recombination M can combine the two alleles to create the fitter genome a lot more quickly. As such, the modifier allele M will become associated with the alleles and spread through hitch-hiking (OTTO and BARTON 1997; ROZE and BARTON 2006). Figure modified from MULLER (1932).

tions are mainly deleterious (which has been confirmed by subsequent studies (EYRE-WALKER and KEIGHTLEY 2007)), and back mutation to restore the wildtype allele is rare. Both conditions will lead to the buildup of deleterious mutations, but initially there will exist a proportion of mutation-free individuals. In finite populations, such a class of individuals will eventually be lost and cannot be recreated. This loss constitutes one 'click' of the ratchet. Over subsequent generations the loss of the least-loaded class will continue, which would also lead to deleterious alleles fixing in the population (CHARLESWORTH and CHARLESWORTH 1997). This process can cause an irreversible degradation of the genome that can drive a population to extinction (LYNCH *et al.* 1993). Sex and recombination are therefore beneficial by recreating genomes having smaller numbers of deleterious mutations, preventing degradation of the genome over time. Subsequent work has shown that only a small amount of recombination is needed to stop the ratchet (CHARLESWORTH *et al.* 1993). Furthermore, deleterious mutations would not build up if compensatory mutations arise, except in very small populations (POON and OTTO 2000; KAISER and CHARLESWORTH 2009).

General selection interference and the 'Hill-Robertson' effect. In the years following MULLER (1964), there was renewed controversy around whether the 'Fisher-Muller' mechanism or Muller's Ratchet could explain the ubiquity of sex, especially in comparisons of finite-populations models against infinite-population models (reviewed in FELSENSTEIN (1974)). The Fisher-Muller hypotheses and Muller's Ratchet were explored and united in a seminal paper by FELSENSTEIN (1974), which explained how the Fisher-Muller hypothesis and Muller's Ratchet are conceptually the same as what he described as 'The Hill-Robertson effect', named after the paper by HILL and ROBERTSON (1966). In it, a mixture of diffusion equations and computer simulations were used to demonstrate that in finite populations, selection and drift creates chance associations between alleles, with negative associations persisting for longer. Genetically, this reflects the reduced effectiveness of selection at a specific allele through chance associ-

ations with a selected linked locus. This process leads to an increase in fitness variance at the focal site, due to the effect of selection acting on linked sites, but a reduction in total fitness variance in the population, usually (but not always) leading to a subsequent decrease in the effective population size, N_e (COMERON *et al.* 2008). FELSENSTEIN (1974) then used simulations to investigate how recombination increases the fixation rate of beneficial mutations (Fisher-Muller hypothesis), and stops the build-up of deleterious mutations (Muller's Ratchet), if different selection coefficients and mutation rates were used.

Further research has investigated various types of selection interference that can be described as a 'Hill-Robertson' effect. Such processes generally lie in one of four main categories, as summarised by CHARLESWORTH *et al.* (2009):

1. *Genetic hitchhiking*. A selective sweep arising in the genome can drag alleles at linked loci with it to fixation, reducing levels of genetic variability around it (MAYNARD SMITH and HAIGH 1974). Neutral alleles are generally affected, although weakly deleterious alleles could also hitchhike in regions of low recombination (HADANY and FELDMAN 2005; HARTFIELD and OTTO 2011; CHUN and FAY 2011), or polymorphism could be lost at linked loci under balancing selection (PECK 1993). Sweeps can also interfere with fixation of new advantageous alleles if they arise at linked sites, as described in the Fisher-Muller hypothesis (BARTON 1995b) (this is also known as 'clonal interference' if acting in clonal organisms (GERRISH and LENSKI 1998)).

2. *Background selection*. Deleterious mutations enter the population via mutation and are generally removed quickly by selection. This mechanism also removes neutral variation around the site of deleterious alleles (CHARLESWORTH *et al.* 1993, 1995; HUDSON and KAPLAN 1995). Background selection can also impede the spread of advantageous alleles (JOHNSON and BARTON 2002; BACHTROG

and GORDO 2004), and allow other deleterious alleles to persist in the population for longer (BARTON 1995b).

3. *Muller's Ratchet*, as described above.

4. *Weak Hill-Robertson effects*. If a large number of linked sites are subject to reversible mutation between advantageous and deleterious states, then linkage can cause deleterious alleles to persist at frequencies above that expected by mutation-selection equilibrium (MCVEAN and CHARLESWORTH 2000; KAISER and CHARLESWORTH 2009).

Because of all these different classifications, most of the classical theories explaining the evolution of sex in finite populations are usually described as special cases of a single theory, in which recombination is beneficial by breaking apart interfering mutations and increasing the response to selection.

'Hill-Robertson' effects selecting for sex and recombination. In a general analysis of the Hill-Robertson effect, BARTON (1995b) showed how recombination between two selected alleles could increase the efficacy of selection acting on both. OTTO and BARTON (1997) then demonstrated that this mechanism can select for increased levels of recombination at a modifier locus, but the increase in frequency of a recombination modifier is only significant if linkage between loci is initially tight, as even a small amount of recombination can greatly increase the efficacy of selection. Further analyses demonstrated how negative linkage disequilibrium arises by chance in finite populations, even if the population started in linkage equilibrium, which leads to selection for increased recombination (BARTON and OTTO 2005). Recombination also helps to accelerate the spread of advantageous alleles over time, after interference is broken down (ROZE and BARTON 2006). However, BARTON and OTTO (2005) and ROZE and BARTON (2006) concluded that although breaking down interference selects for higher

levels of recombination, the effect was only strong in small populations and could not explain the evolution of costly sex in large populations.

This view has changed in recent years, with the finding that breaking down selection interference can select for increased levels of recombination in larger populations, if it acts over multiple linked loci. OTTO and BARTON (2001) showed that if acting over three loci experiencing directional selection, a modifier was not favoured in populations of $N > 10,000$ chromosomes unless synergistic epistasis was present between loci (which also creates negative associations that selects for recombination; see 'Mutational Deterministic' section). However, in 11-locus simulations, a modifier was selected for in populations larger than 10,000 individuals. Extending this, ILES *et al.* (2003) demonstrated that stronger selection for recombination modifiers arose in very large, finite populations (the largest population investigated consisted of $100,000$ haploid individuals), if individuals contained more loci experiencing directional selection. KEIGHTLEY and OTTO (2006) then showed that recombination could be very strongly selected for in large populations consisting of individuals subject to recurrent deleterious mutation across multiple, linked loci. It was also shown that these large advantages to recombination could potentially overcome a twofold cost of sex, but only if modifier genes increased the frequency of sex to low levels.

The studies mentioned above mainly considered selection for recombination, if populations are subject to deleterious mutation (except ILES *et al.* (2003), where mutations were solely advantageous). However, such work did not consider the effects of both advantageous and deleterious mutations acting together. PECK (1994), CHARLESWORTH (1994) and PECK *et al.* (1997) demonstrated how sex can move novel advantageous alleles away from deleterious genetic backgrounds, increasing their fixation probability, leading to a higher population mean fitness. Therefore the presence of both advantageous and deleterious mutation can offer additional selection for recombination. This was demonstrated by HARTFIELD *et al.* (2010) (see also Chapter 3), whose simulations

showed greater selection acting on a recombination modifier in populations subject to recurrent advantageous and deleterious mutation at multiple loci, compared to populations exposed to deleterious mutation only.

The effect of spatial structure. PECK *et al.* (1999) and SALATHÉ *et al.* (2006) demonstrated that costly sex can be maintained against asexual invasion if there is sufficient population subdivision. This result is a consequence of asexuals accumulating excessive deleterious mutations, since population subdivision increases the time needed for an invading asexual to fix in the entire population. Figure 1.3 shows a schematic of this process. SALATHÉ *et al.* (2006) also noted that population subdivision disfavours asexuals since the smaller within-deme population sizes increase the rate of Muller's Ratchet (GESSLER 1995; HIGGINS and LYNCH 2001). The benefits of sex also increase with population size, since this further slows down the spread of an asexual (SALATHÉ *et al.* 2006). MARTIN *et al.* (2006) demonstrated that subpopulations can maintain polymorphism, increasing selection on a recombination modifier even in very large overall population sizes. However, HARTFIELD *et al.* (2012) (see also Chapter 5) explored asexual invasion into a structured population subject to deleterious and advantageous mutation, with various levels of population subdivision. It was found that large populations ($N = 40,000$ overall) needed fairly high levels of population subdivision, as measured by F_{st} (≈ 0.3 was the lowest level found), to maintain sex with a twofold cost. The critical level of F_{st} needed to maintain sex decreased if the number of subpopulations increased. Overall, the greatest advantage to sex arose if demes were arranged in a stepping-stone formation spread over a large number of demes, with individuals subject to both advantageous and deleterious mutation. This finding suggests that population subdivision could possibly maintain costly sex in very large populations spread over a large number of demes, but such a scenario is currently computationally intractable.

The effect of diploidy. All the above studies considered haploid individuals, but

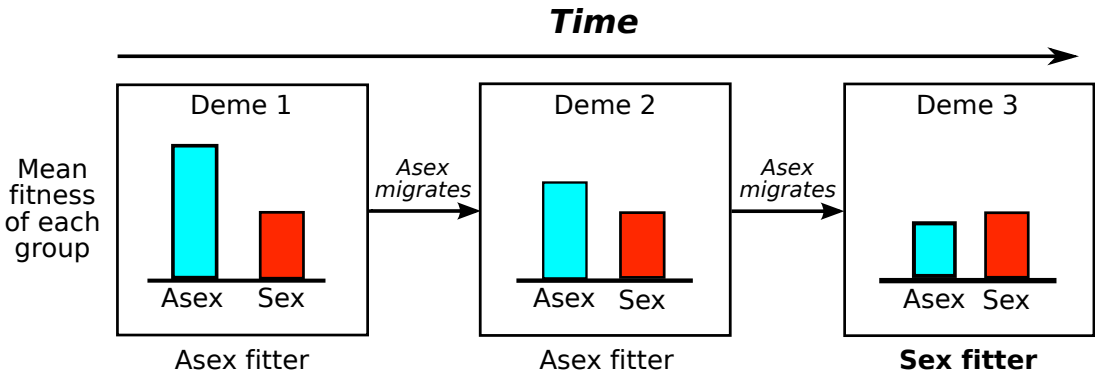


Figure 1.3: The maintenance of sex in a subdivided population. Initially an asexual has a twofold advantage and fixes in the first deme. However, when it migrates to a second deme its advantage would be lower due to deleterious mutation accumulation. Given enough time and demes to transfer to, asexuals would eventually have a lower mean fitness than sexuals, so sexuals eventually outcompete asexuals. Only three demes are showed here for brevity, but the argument can be applied to any number of demes, given a low enough migration rate.

most animal and plant species, as well as other higher eukaryotes in nature are diploid. Not only do sexual diploids undergo crossing-over, which alters the arrangement of alleles on different loci at the same chromosome, but also segregation, which alters the associations within a specific locus. In infinite populations, segregation can cause sexuals to obtain a higher mean fitness through restoring the fittest homozygous genotype (CHASNOV 2000; OTTO 2003), and aid the fixation of advantageous alleles in diploid populations with intermediate dominance ($h = 1/2$) (KIRKPATRICK and JENKINS 1989). Furthermore, sex is selected for under a wider range of deleterious selective strengths and dominance values in inbred populations, as inbreeding is more likely to create unfit homozygotes that are subsequently purged by selection (AGRAWAL and CHASNOV 2001; OTTO 2003).

Recent models have incorporated finite population size effects, where selection interference also arises between loci. HAAG and ROZE (2007) showed that at a single locus subject to deleterious mutation, with recessive heterozygote mutants ($h < 1/2$),

small sexual populations (generally less than $N = 10,000$) have a lower fitness load compared to asexuals, where the fitness load is defined as the difference between the population's mean fitness and its maximum possible value. The load is reduced in sexuals because segregation breaks down negative associations that arise due to drift within loci. These negative associations are more prevalent in subdivided populations, generating higher loads in asexuals relative to sexual populations.

However, ROZE (2009) and ROZE and MICHOD (2010) showed that recombination and sex modifiers in finite populations are selected *against* if deleterious mutants are partially recessive ($h < 0.2$), because recombination tends to break correlations in heterozygosity across multiple, linked loci, and thus reduce the frequency of genotypes that are heterozygous at multiple loci, which have the highest fitness. This result makes it harder to explain whether breaking apart selection interference could explain the ubiquity of sex, since there is evidence that most deleterious alleles are partially recessive in nature (SIMMONS and CROW 1977).

Experimental evidence for selection for sex breaking down selection interference. Several experimental studies have demonstrated that sex can increase the efficacy of selection by breaking down interference between loci. I shall only focus on those demonstrating that breaking apart selection interference creates a benefit for sexuals populations compared to asexuals. There are many other studies demonstrating that recombination increases the efficacy of selection at the genetic level; these are discussed further in CHARLESWORTH *et al.* (2009) and references therein.

MALMBERG (1977) verified the Fisher-Muller hypothesis using the bacteriophage T4, by observing that cells that interchanged strands with others have a higher rate of adaptation, compared to those with no recombination. A similar finding was observed in *D. melanogaster* by RICE and CHIPPINDALE (2001), and in *Chlamydomonas reinhardtii* where sexuals were better able to overcome the effects of clonal interference, leading to an accelerated adaptation (COLEGRAVE 2002). However, whilst those stud-

ies showed that sex is advantageous through accelerating adaptation, ZEYL and BELL (1997) found that sexual strains of yeast only increased mean fitness in environments to which populations were already adapted, as opposed to strains placed in novel environments. This finding suggests that sex is maintained predominantly through preventing the buildup of deleterious mutations via Muller's Ratchet. POON and CHAO (2004) used population bottlenecking in a system of the RNA bacteriophage $\Phi 6$ in order to mimic the effects of genetic drift, and observed that asexuals have a lower response to selection compared to sexuals. Similarly, sexuals strains of yeast in a stressed environment had a higher fitness compared to asexuals, and a higher variance in fitness, again increasing the response to selection, in line with Weismann's hypothesis (GODDARD *et al.* 2005). MORRAN *et al.* (2009) showed that *Caenorhabditis elegans* evolved outcrossing not only to increase the rate of adaptation, but also to prevent deleterious mutational meltdown. Finally, from molecular evolution analyses, it was determined that obligate asexual species of *Daphnia pulex* microcrustacean (PALAND and LYNCH 2006) and asexual lineages of *Potamopyrgus antipodarum* snails (NEIMAN *et al.* 2010) were found to accumulate deleterious mutations at an accelerated rate in mitochondrial genomes, as measured using polymorphism to divergence ratios at noncoding and synonymous sites.

1.3 Sex to resist parasites; escaping the 'Red Queen'

The coronation of the Red Queen. During the 1970s, the hypothesis that sex evolved to break down selection interference was subject to much criticism (as reviewed in MAYNARD SMITH (1978) and JAENIKE (1978)). At the time it was seen as a 'group selectionist' argument; that is, sex evolves due to the benefit it offers to a group of individuals, as opposed to the benefits offered to modifier genes that increased levels of sex. It was also unclear at the time whether there was a sufficiently long-term benefit to breaking down selection interference that could explain the maintenance of sex.

Because of this, JAENIKE (1978) proposed, using a verbal model, that sex could be beneficial through creating rare genotypes, which can resist parasites that have adapted to infect individuals with a specific genotype.

This hypothesis gained prominence after HAMILTON (1980) formalised Jaenike's argument and demonstrated that, with fluctuating selection, sexuals could gain a twofold fitness advantage over asexuals, if the recombination rate and fecundity of sexuals was high enough. These prerequisites enabled the creation of a higher frequency of rare genotypes that are resistant to parasites. However, by more explicitly considering the epidemiological dynamics of parasitic infection (such as pathogen reproductive and infection rates), MAY and ANDERSON (1983) showed that sex only had a twofold fitness advantage if infection was nearly lethal to the host. HAMILTON *et al.* (1990) later demonstrated that given a high enough number of loci that determined the host's infection susceptibility, and if individuals were subject to rank-order truncation selection (that is, the least fit are automatically killed off), then costly sex could fix in a population. Note however that this model assumes individuals consisting of multiple linked loci subject to selection, and truncation selection that creates synergistic epistasis between deleterious loci. Sex could therefore confer additional benefits through breaking apart selection interference, or restoring fitness variance according to the 'mutational deterministic' model (see next section).

This basic model suffered from limitations, specifically with regards to whether the creation of rare genotypes could select for modifiers genes for increased recombination. LADLE *et al.* (1993) extended Hamilton's model to a subdivided population, and demonstrated that if there exists large discrepancies between host and parasite migration rates then sex would be selected against. This is because migration restores genotypes from other regions that would otherwise be rare in that deme, removing the benefits to sex. With regards to sex in fluctuating environments, CHARLESWORTH (1976) and BARTON (1995a) showed that linkage disequilibrium needs to change sign rapidly (every two

to five generations) for recombination to be favoured in infinite populations. Furthermore, OTTO and NUISMER (2004) showed that under a variety of species interactions, increased levels of recombination would be disfavoured because adapted gene combinations would be broken apart, unless the fitness reduction due to parasitic infection was strong.

Expansion of Red Queen models. Subsequent investigations of the Red Queen hypothesis aimed to answer one of two main questions. Firstly, what processes select for increased sex and recombination under this mechanism? Secondly, are there any ways in which the basic model can be extended so that host-parasite interactions selects for sex under a wider range of biologically-realistic conditions?

Concerning to the first question, PETERS and LIVELY (1999) determined that antagonistic coevolution leads to fluctuating linkage disequilibrium and epistasis in parasites and hosts. This leads to selection for recombination modifiers through the creation of rare genotypes that confer higher fitness. Conversely, there was little advantage to recombination through increasing additive genetic variance in fitness, which improves an individual allele's response to directional selection. This result was later formalised by GANDON and OTTO (2007), who also found that increased parasitic virulence causes more rapid fluctuations in host and parasite genotypes, which is the conditions that favour higher rates of recombination. This finding could be used to determine whether advantages to sex observed in field and experimental studies arise due to Red Queen dynamics, or instead through breaking apart selection interference (BARTON 2010; penultimate paragraph). This point is especially of interest, since BARTON and OTTO (2005) demonstrated that selection interference is present at loci subject to fluctuating selection in finite populations, and recombination can be beneficial through breaking it down. PETERS and LIVELY (2007) and SALATHÉ *et al.* (2009) later determined that recombination modifiers spread mainly due to a 'delayed short-term benefit', which is the advantage to recombination through creating rare genotypes that have maximum pos-

sible fitness in the generations immediately after their creation.

The basic Red Queen models have been extended in numerous ways. LYTHGOE (2000) demonstrated that if parasites attack a vertebrate with an adaptive immune system, sex is selected for within parasites, however such advantages could not overcome a twofold cost. Similarly, LIVELY (2010a) modified MAY AND ANDERSON'S (1983) model, so that the infection rate is proportional to the number of infected hosts, leading to sexual and asexual hosts coexisting over time. Other studies found scenarios where sex could be maintained under a broader range of conditions, such as if 'similarity selection' was assumed (that is, the offspring that a certain parasite is likely to infect is genetically similar to the host's parent) (AGRAWAL 2006); strong selection against non-infecting parasites relaxes the need for parasitic infection to be strongly deleterious to hosts (SALATHÉ *et al.* 2008b); similarly, if virulence is density-dependent and the population death rate is low, then parasitic infection can maintain sex, even if it is weakly virulent (LIVELY 2009, 2010b). However, AGRAWAL (2009a) found that conditions favouring evolution of recombination in haploid Red Queen models did not generally cause equally strong selection for sex in diploids. MOSTOWY *et al.* (2010) also found different dynamics if multiple parasites could infect an individual; sex was generally selected against if simultaneous infection of a host was common, because this broke down fluctuating linkage disequilibrium.

Empirical evidence for the Red Queen hypothesis. Although existing theoretical work suggests that the Red Queen hypothesis only selects for sex under specific circumstances, there exists a wide body of empirical studies that shows parasite interactions selecting for increased levels of sex in nature, based on field studies and recent experimental work.

The fundamental prediction of the Red Queen hypothesis - that exposure to parasitic infections maintains sexual reproduction - has been directly tested and observed in both field and laboratory studies. JOKELA *et al.* (2009) clearly showed that pre-

viously common asexual clones of *Potamopyrgus antipodarum* snails were driven to extinction within a few years, whilst sexuals remained at a high frequency. Common clones were also more susceptible to sympatric *Microphallus* sp. parasites (those that arise in the same region as the asexual population under investigation), leading to negative frequency-dependent selection. KING *et al.* (2009) provided the first experimental evidence of Red Queen dynamics maintaining sex in *P. antipodarum*. In a direct experimental test of the Red Queen hypothesis, MORRAN *et al.* (2009) showed that *C. elegans* evolved higher levels of outcrossing when exposed to the bacterial pathogen *Serratia marcescens*. MORRAN *et al.* (2011) subsequently demonstrated that 'sexual' *C. elegans*, which were genetically manipulated to be obligate outcrossers, became fixed very quickly if populations were constantly exposed to coevolving strains of the pathogen, indicating that the presence of coevolving parasites directly selected for sex. Other studies by LIVELY (1987) and KUMPULAINEN *et al.* (2004) demonstrated the existence of a positive correlation between parasitic infection and the frequency of sexuals. In lab experiments, FISCHER and SCHMID-HEMPEL (2005) observed significantly increased levels of recombination in *Tribolium castaneum* beetles subject to infection by the *Nosema whitei* parasite.

Field studies and lab experiments have also verified specific assumptions and predictions of the Red Queen hypothesis. These assumptions include whether parasitic infection is frequency-dependent and infection disproportionally affected asexuals; whether asexuals were derived from sexual ancestors (as assumed in the model of HAMILTON *et al.* (1990)); and also determining the selective disadvantage of noninfected parasites. DYBDAHL and LIVELY (1995) demonstrated that asexual and sexual species of *P. antipodarum* coexist in the same area, with asexuals evolving from sexual species, verifying the assumption that asexual clones are derived from and compete with local sexual populations. Such asexuals mostly diverged 20,000-70,000 years ago, with a few ancient asexuals also present with divergence times of 500,000 years (NEIMAN *et al.* 2005).

KING *et al.* (2011a) showed that increased rates of infection also promoted clonal diversity. and trematode parasites die if they fail to infect the host, offering evidence for strong selection against parasites (KING *et al.* 2011b), which is needed to maintain sex under the Red Queen hypothesis (SALATHÉ *et al.* 2008b). KOSKELLA and LIVELY (2009) verified the prediction of the Red Queen hypothesis that common host genotypes should reduce in frequency over time, due to parasitic infection.

1.4 The 'Mutational Deterministic' hypothesis

In the 'Mutational Deterministic' hypothesis, negative disequilibrium persists in infinite populations subject to deleterious mutation, if deleterious mutants act synergistically. That is, a collection of deleterious mutants will cause a larger detriment to an individual's fitness than expected if they acted independently. This leads to the evolution of increased levels of recombination, which can overcome a twofold cost of sex if the deleterious mutation rate is high enough. However, it will be seen that the clear yet strict conditions needed to maintain obligate sex under this model are not found to be widespread in nature. As a consequence the hypothesis is losing favour as an explanation of the evolution of sex and recombination.

The mutational deterministic hypothesis attracted attention when several theoretical papers appeared demonstrating that recombination can be selected for in infinite populations. Previous models had shown that recombination is selected against because it breaks apart the fittest genotype (NEI 1967; OTTO 2009). The necessary condition needed for recombination to be advantageous is the presence of negative linkage disequilibrium between loci, which can be generated by synergistic epistasis (FELDMAN *et al.* 1980). Recombination places more deleterious mutants together in the same genome, which causes a larger fitness reduction than if each mutant acted individually, due to the presence of epistasis. These individuals are then more likely to die, purging more

deleterious alleles from the population (KONDRASHOV 1982). This process leads to a highly reduced fitness load in sexuals, whilst the load in asexuals is always equal to $1 - e^{-U}$ at mutation-selection equilibrium, for genome-wide deleterious mutation rate U, irrespective of the distribution of fitness effects (KIMURA and MARUYAMA 1966). Furthermore, if the genomic mutation rate U is greater then one, then the benefit is large enough to overcome the two-fold cost of sex (KONDRASHOV 1993).

CHARLESWORTH (1990) subsequently investigated in detail the role played by epistasis in selecting for sex, and showed that the benefits to recombination increased with more chromosomes, and longer map length per chromosome. BARTON (1995a) later quantified what range of epistasis would select for increased recombination (see OTTO and LENORMAND (2002) for a summary of the main results). It was shown that in order for epistasis to create negative associations between loci that persist over time, it has to be weak and negative. Specifically, under a quasi linkage equilibrium (QLE) scheme (where selection and epistasis are weak relative to the recombination rate; see KIMURA (1965); BARTON and TURELLI (1991); KIRKPATRICK *et al.* (2002)), where the recombination modifier has a small effect on overall recombination rate ($\delta p \ll 1$), then the rate of increase of a recombination modifier is equal to:

$$\frac{\delta p}{\rho_{MAB}} D(\lambda - \epsilon) \tag{1.1}$$

where δp is the increase in recombination caused by the modifier; ρ_{MAB} the recombination rate across the set of three loci (modifier locus M and selected loci A, B); D the linkage disequilibrium between selected loci; ϵ the degree of epistasis, and λ is a compound parameter that takes the form:

$$\lambda = -s_a s_b \left[\frac{1}{\rho_{MA}} + \frac{1}{\rho_{MB}} - 1 \right] \tag{1.2}$$

Here s_a, s_b are the selection coefficients of the deleterious alleles at selected loci A

and B. Equation 1.1 encapsulates both the short-term effect of recombination through breaking apart favoured combinations of alleles, and creating a long-term increase in the population mean fitness. Since $D < 0$ is needed for recombination to be selected for (otherwise recombination is disadvantageous through breaking apart adapted collections of alleles), Equation 1.1 shows that epistasis should lie in the range $\lambda < \epsilon < 0$. Figure 1.4 outlines a schematic of the parameter space needed to select for recombination. OTTO and FELDMAN (1997) later verified this analysis, and also showed that initial levels of linkage have to be tight in order for increased recombination to evolve.

Testing the Mutational Deterministic hypothesis. The above analysis makes clear that in order for sex to be widespread, the genomic mutation rate should be high (generally greater than or equal to one), and synergistic epistasis needs to exist between deleterious mutants. Because of these simple predictions, and its focus on deleterious mutants (which are prevalent in nature), it was, until recently, an appealing explanation for widespread costly sex.

More is now known about the deleterious mutation rates for several species. Some do seem to have genomic deleterious mutations rates greater than one, such as Drosophila (HAAG-LIAUTARD *et al.* 2007; KEIGHTLEY *et al.* 2009), *C. elegans* (DENVER *et al.* 2004) and hominids (EÖRY *et al.* 2010). Studies generally show equivocal evidence as to whether a deleterious mutation rate greater than one is common in coding regions (KEIGHTLEY and EYRE-WALKER 2000; BAER *et al.* 2007). However, since non-coding regions are also subject to deleterious mutation (EÖRY *et al.* 2010), it is reasonable to assume that U can easily exceed one in higher eukaryotes. More importantly, widespread net epistasis has not been found in nature; a large variance in epistasis is generally found in organisms, which disfavours the evolution of increased recombination (OTTO and FELDMAN 1997). See KOUYOS *et al.* (2007) for a review of data, as well as a discussion on the limitations of these studies, such as the fact that only large-effect epistatic interactions can only be reasonably detected.

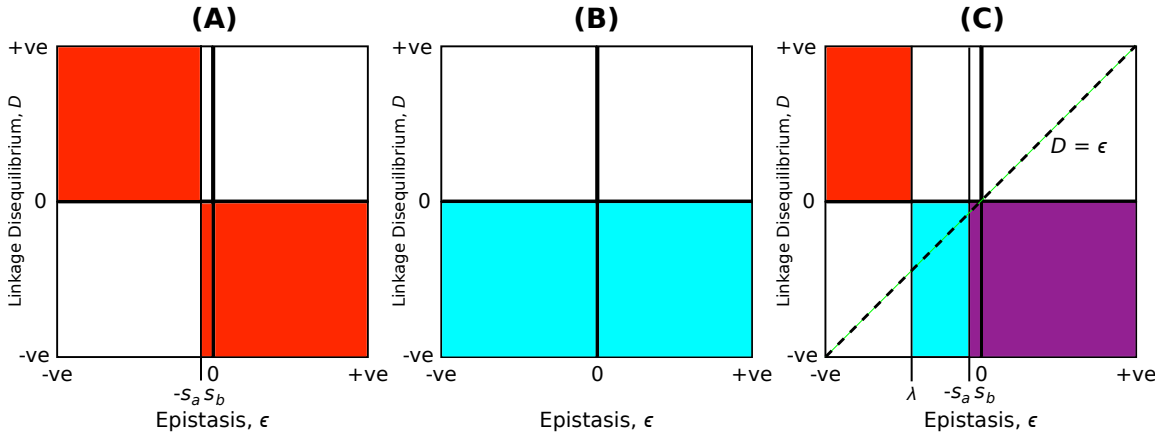


Figure 1.4: Explaining the parameter space under which recombination is favoured according to the 'Mutational deterministic' model. The $+ve$ and $-ve$ terms on the axis indicate where the epistasis and linkage disequilibrium parameters are positive and negative, respectively. (a) There is a short-term advantage if the mean fitness of 'extreme' genotypes (AA and aa) is greater than that of 'intermediate' genotypes (Aa and aA); recombination is beneficial if it breaks up intermediate genotypes in this case, which arises (approximately) if D and ϵ have opposite signs (red sections). (b) Recombination offers a long-term advantage through increasing the genetic variance in the population, which arises if intermediate types are under-represented; that is, if $D < 0$ (blue sections). (c) A modifier is therefore selected for if the short-term advantage outweighs the long-term disadvantage (red); if the long-term advantage outweighs short-term disadvantage (blue); or if both short-term and long-term effects are beneficial (purple). In an infinite model in a homogeneous environment, disequilibrium is only formed through epistasis, thus recombination can only evolve when the $D = \epsilon$ line overlaps the key areas, which arises if $\lambda < \epsilon < 0$. Figure adapted from LENORMAND and OTTO (2000); OTTO and LENORMAND (2002).

It should be noted, however, that a high mutation rate and synergistic epistasis is only needed for sex to overcome a twofold cost. Deleterious mutations that act synergistically, but arise at lower rates, can still cause higher rates of recombination to evolve (KOUYOS *et al.* 2007). Sex could have a twofold advantage if mutational deterministic processes are combined with other mechanisms, such as fluctuating selection (WEST *et al.* 1999), or finite-population effects such as Muller's Ratchet (HOWARD and LIVELY 1998, 2002). However, the current general consensus is that the conditions required to fulfil the mutational deterministic hypothesis are not widespread in nature, and other mechanisms can more easily explain the production of negative associations between loci that are needed to select for sex and recombination (KOUYOS *et al.* 2007).

1.5 Other hypotheses on the evolution of sex and recombination

The three hypotheses discussed so far - selection interference, 'Red Queen' dynamics and the 'Mutational Deterministic' model - are the most prominent mechanisms proposed to explain the maintenance of costly sex. Nevertheless, alternative, less prominent ideas have been proposed that can aid selection for sex.

An idea that has been recently explored extensively is the idea of 'Fitness-Associated Recombination' (FAR), and the related mechanism of 'Fitness-Associated Sex' (FAS). This is the idea that if populations are subject to environmental stress, the level of recombination and sex increases in individuals with lower fitnesses, so as to restore the fittest genotype (HADANY and OTTO 2007). Recombination and sex would then be associated with the fittest individuals, causing modifiers coding for fitness-associated recombination and sex to spread. Research into this mechanism was motivated by numerous experimental studies showing that organisms are more likely to evolve higher levels of recombination if subject to fitness stresses (HADANY and BEKER (2003) and

references therein).

REDFIELD (1988) first demonstrated that FAR is beneficial in a model consisting of less-fit bacteria individuals undergo transformation, whilst the fittest individuals reproduced asexually. This leads to an increase in the mean fitness of the population. GESSLER and XU (2000) subsequently showed, using computer simulation, that FAR is selected for in populations at linkage equilibrium, where recombination modifiers that are not fitness-dependent would be selectively neutral. HADANY and BEKER (2003) then demonstrated analytically that populations subject to FAR have a higher mean fitness, compared to those subject to uniform recombination rates (UR). This result holds either in infinite populations with one or two loci, if selection at each locus is stronger than the deleterious mutation rate, as FAR was more able to create positive associations between alleles in lower-fitness individuals, compared to UR. This scenario resulted in FAR populations having a higher mean fitness compared to equivalent UR populations. AGRAWAL *et al.* (2005) subsequently found that whilst FAR is advantageous in haploids, the benefits did not extend to diploids (except if *cis-trans* effects were present), since recombination was likely to break apart advantageous associations between the recombination modifier and a beneficial allele in heterozygotes.

However, Fitness-Associated Sex (FAS) could be beneficial in diploids due to segregation. HADANY and OTTO (2007) found that FAS could evolve if the costs of sex were not too high. Although the dynamics of such a modifier were complex, one steady-state that was found was the 'extreme' case where the fittest class of individuals were purely asexual, with others undergoing some degree of FAS. Subsequent research showed that FAS accelerated adaptation in haploids and diploids, if there were some cost of sex (so as to activate the fitness-dependent sex mechanism), and if heterozygotes engaged in sex more often than expected, compared to the multiplicative rates of sex in homozygotes (HADANY and OTTO 2009).

Another recently-discussed hypothesis is that sex and recombination are favoured

in populations spread over heterogeneous environments, where different alleles at the same locus have different, environmental-dependent selective effects. In such a situation, recombination can be favoured in an infinite population of haploids if there is a wider range of epistasis values between the selected loci compared to homogeneous populations, depending on the covariance in selection coefficients between loci across environments (LENORMAND and OTTO 2000). Similar to the FAS mechanism, AGRAWAL (2009b) found an advantage to sex in spatially heterogeneous environments, if the fitness of heterozygotes is greater than the mean fitness of the two homozygotes at a locus (Figure 1.5). This is because migration causes an excess of homozygotes to be introduced into each population, so sex is advantageous through creating the fitter heterozygote state.

Unfortunately there is limited experimental evidence as to whether fitness-associated mechanisms or heterogeneous environments select for increased sex in nature. SCHOUSTRA *et al.* (2010) found evidence for FAS emerging in the fungus *Aspergillus nidulans*, and suggested that it could represent "a feasible first step in the evolution of sexual reproduction". A recent study found that increased levels of sex emerge in the monogonont rotifer *Brachionus calyciflorus*, if it was transferred between two heterogeneous environments (BECKS and AGRAWAL 2010). No evidence was found that such sex was fitness-dependent. It was also asserted that drift effects did not select for higher levels of sex, since there were no differences in the frequency of sexuals if the population size increased, and sexuals did not have a higher variance in fitness across the entire population (BECKS and AGRAWAL 2011).

1.6 Conclusions and direction of PhD research

Despite many years of research and study, a thorough explanation as to how costly sex is widespread in nature is still elusive. Nevertheless, as the above review shows, there are

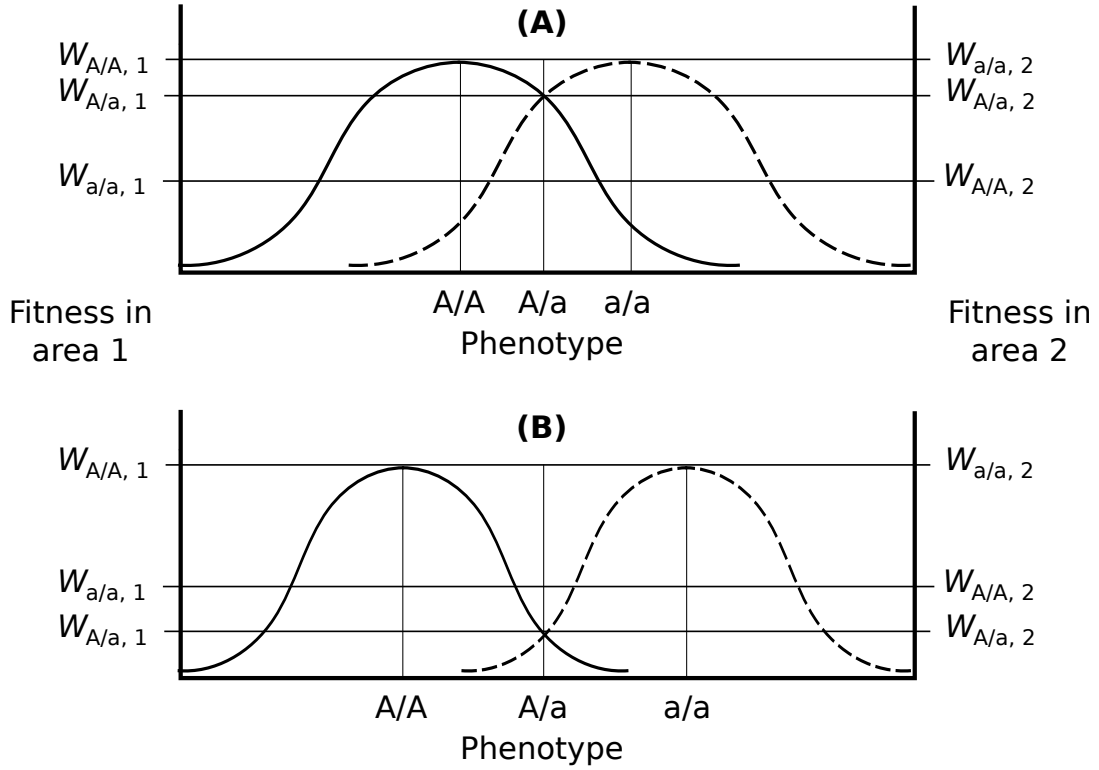


Figure 1.5: An example of where sex would be favoured in diploids in spatially heterogeneous environments. (a) The diploid locus is under stabilising selection; since the fitness of the heterozygote is greater than the mean of the homozygotes, sex would be advantageous as it would form more heterozygotes from the excess homozygotes created by migration. (b) If the optima lie too far apart then heterozygotes would be disadvantageous, so sex would be selected against. Figure adapted from AGRAWAL (2009b).

no shortage of theories as to how sex and recombination could evolve and be maintained in nature. Years of theoretical study means that the main genetic mechanisms driving sex and recombination are now well-understood. There has also been an increase in the use of experimental systems and data from field studies to test individual hypotheses, and to ascertain what mechanisms (selection interference, host-parasite interactions, epistasis, spatial heterogeneity) select for sex in the wild.

For my PhD, I aimed to extend the study of KEIGHTLEY and OTTO (2006), and further investigate whether the breaking apart of selection interference across multiple linked loci still favours increased levels of recombination if different types of mutation were present; for example, if a small amount of advantageous mutation occurred alongside deleterious mutation. It was also of interest to determine whether the advantages to recombination found by KEIGHTLEY and OTTO (2006) were also large enough to explain the maintenance of costly sex, especially if such a sexual population was subdivided (as previously investigated by PECK *et al.* (1999); SALATHÉ *et al.* (2006)). Furthermore, it would also be of interest to investigate analytically (as opposed to just using computer simulations) how specific mechanisms introduced into these models, such as the effect of advantageous mutation and population subdivision, affects selection for recombination and costly sex.

This thesis will therefore contain the following chapters:

- **Chapter 2: Simulation of selection on two linked loci, and investigation into the Hill-Robertson effect.** An introductory data chapter, recreating the simulation model from HILL and ROBERTSON (1966), demonstrating how selection interference arises in a two-locus system.

- **Chapter 3: The role of advantageous mutations in enhancing the evolution of a recombination modifier.** Extension of the model presented in KEIGHTLEY and OTTO (2006), investigating how the presence of a small fraction of advantageous mutations enhances selection acting on a recombination modifier, compared to

cases where mutation is just deleterious. This effect is studied further by measuring the underlying variance in fitness, linkage disequilibrium, selection acting on the modifier, and reduction in effective population size (HARTFIELD *et al.* 2010).

- **Chapter 4: Recombination and hitchhiking of deleterious alleles.** One advantage of recombination as observed in Chapter 3 is that it prevents weak deleterious alleles from hitchhiking with advantageous mutants, which would otherwise fix if there was no recombination between the two loci. Here an analytical model is used to investigate the fixation probability of such deleterious alleles, given the strength of selection acting at each site, and the recombination rate between them (HARTFIELD and OTTO 2011).

- **Chapter 5: Can weak population structure protects sexual populations from asexual invasion?** If recombination acting over multiple loci subject to advantageous and deleterious mutation is strongly advantageous, could this benefit overcome a twofold cost of sex in structured populations? I run simulations to investigate whether moderately-large populations consisting of a few tens of thousands of individuals can maintain costly sex with this mechanism, with modest levels of population structure (as measured using F_{st}).

- **Chapter 6: A general framework for estimating the fixation time of an advantageous allele in stepping-stone models.** In order to analytically investigate how the fixation time of an asexual invader increases as it invades a structured population, methods need to be developed in order to determine the effect that population structure has on the spread of sweeping mutants. Here, a method is laid out that can be used to measure the fixation time of an advantageous allele in a subdivided population. This method can then be applied to a variety of stepping-stone models.

- **Chapter 7: Conclusions and further research.** A discussion of the findings

obtained over the course of my PhD, and the limitations of the models used; and future directions that research into the evolution and maintenance of sex should focus on.

Chapter 2

Simulation of selection on two linked loci, and investigation into the Hill-Robertson effect

Chapter abstract

The 'Hill-Robertson' effect, where selection and drift acting on linked loci interferes with and reduces the strength of selection acting at each, is an important phenomenon in evolutionary genetics. It can create negative associations between loci that persist over time. Recombination can then be advantageous by breaking apart such interference and aiding the fixation of advantageous mutations. In this chapter, the simulation model outlined in HILL and ROBERTSON (1966) is recreated, in order to determine how recombination affects the fixation probability of individual beneficial alleles in a two-locus system, as well as the fixation of the fittest genotype. These results are analysed in light of analytical and simulation work that followed HILL and ROBERTSON (1966).

2.1 Introduction

The study by HILL and ROBERTSON (1966) is a seminal paper in computational biology, which demonstrated how recombination can be beneficial through bringing together advantageous alleles that arise in different genomes, in line with the Fisher-Muller hypothesis (FISHER 1930; MULLER 1932). This is achieved through breaking down negative linkage disequilibrium that arises by the joint action of selection and drift, leading to the recreation the fittest genotypes (FELSENSTEIN 1974). This process has had a major impact on modern theoretical biology, especially in regard to the evolution of sex and recombination (KEIGHTLEY and OTTO 2006; COMERON *et al.* 2008), and the levels of diversity around selected sites in the genome (CHARLESWORTH *et al.* 2009).

In order to understand this effect further, the simulation outlined in HILL and ROBERTSON (1966) is recreated. This simulation is used to determine how selection and drift in small populations impedes the fixation probability of new alleles, and also how it affects the fixation probability of the fittest possible genotype in the population. This is investigated further by altering the initial frequencies of each selected allele, in order to determine when selection interference would be maximised. I also change the selection strength of each allele, to observe what effect this has on impeding selection at linked loci. Such results are understood more easily in the light of subsequent mathematical and computational analysis of the Hill-Robertson effect.

2.2 Methods

This simulation was based on the methods outlined in HILL and ROBERTSON (1966). Whilst their model initially considered selection acting on two biallelic loci, it was subsequently modified so as to consider these loci affecting two quantitative traits, in order to quantify the effects of linkage on artificial selection response. The model presented

here will differ slightly from that of HILL and ROBERTSON (1966) by only considering selection acting on the alleles at each locus, as opposed to a quantitative model.

A finite diploid population is considered, consisting of $2N$ chromosomes each containing two biallelic loci. The fittest allele at each locus is denoted by A and B, and the unfit alleles by a and b. The fitness differences of the alleles are s_1 and s_2 respectively if present as a homozygote, and half that when present as a heterozygote. An additive model is assumed, so, for example, the fitness of the ab genotype is $(1 - s_1/2 - s_2/2)$. The frequencies of alleles A and B are defined as p and q respectively, with $1 - p$ and $1 - q$ denoting the frequencies of a and b. Initial frequencies p_0 and q_0 are defined at the beginning of the simulation. The frequencies of the four possible haplotypes, AB, Ab, aB and ab, are denoted by f_1, f_2, f_3, and f_4, respectively. Initially, each genotype is at linkage equilibrium; so, for example, the initial frequency of AB is $p_0 q_0$.

The population then goes through a replication cycle consisting of selection, recombination, then reproduction. As in HILL and ROBERTSON (1966), the change in allele frequencies caused by selection and recombination are determined using recursion equations; in this model, these are:

$$f_1' = (f_1 - (1/2)f_1(s_1(1-p) + s_2(1-q)) - \delta D)/\overline{w} \tag{2.1}$$

$$f_2' = (f_2 - (1/2)f_2(s_1(1-p) + s_2(2-q)) + \delta D)/\overline{w} \tag{2.2}$$

$$f_3' = (f_3 - (1/2)f_3(s_1(2-p) + s_2(1-q)) + \delta D)/\overline{w} \tag{2.3}$$

$$f_4' = (f_4 - (1/2)f_4(s_1(2-p) + s_2(2-q)) - \delta D)/\overline{w} \tag{2.4}$$

Here, δD is the change in linkage disequilibrium, defined by $\delta D = rD(1 - s_1/2 - s_2/2)$, and $\overline{w}$ is the mean fitness of the population, and $D = f_1 f_4 - f_2 f_3$ is the standard measure of linkage disequilibrium. The derivation of these equations are given in Appendix 2.A. Note that the change in linkage disequilibrium is altered both by the re-

combination rate r, and selection acting on each locus (denoted by the $(1-s_1/2-s_2/2)$ terms, which is the fitness of the double heterozygotes AB/ab and Ab/aB). The recombination term r, which gives the rate of crossovers occurring between the two selected loci, reduces the magnitude of linkage disequilibrium formed by drift effects, as chance associations between loci are broken down. Comparison of δD with equations 2.1 - 2.4 demonstrates how changing linkage disequilibrium in this manner will benefit the 'extreme' genotypes AB and ab when negative linkage disequilibrium dominates, and the 'intermediate' genotypes Ab and aB when positive linkage disequilibrium dominates. The $(1 - s_1/2 - s_2/2)$ term in δD describe how selection reduces the magnitude of disequilibrium, so chance associations formed by drift do not persist over long periods of time.

After the frequencies of each genotype are changed, $2N$ gametes are chosen at random to form the new generation, with the probability of a new genotype picked based on the current frequency of each genotype. This is the same method as used in HILL and ROBERTSON (1966). $2N$ uniform $[0, 1]$ random numbers are generated, and an AB genotype is created if a random number $X \leq f_1$; Ab created if $f_1 < X \leq f_1 + f_2$; aB chosen if $f_1 + f_2 < X \leq f_1 + f_2 + f_3$, and ab if $f_1 + f_2 + f_3 < X \leq f_1 + f_2 + f_3 + f_4$. This method therefore introduces sampling error into the model, which is needed to produce negative linkage disequilibrium that subsequently reduces the efficacy of selection maintaining the fittest genotypes.

This completes one generation of selection, recombination and reproduction. In each simulation run, the reproduction cycle continued in this way for $6.25N$ generations, or until one of the genotypes fixed in the population. Each simulation was repeated 400 times so that average values can be calculated. The code for a single simulation (time of constant selection and reproduction) was programmed in C (with random numbers provided by algorithms included as part of the GNU scientific library), whilst the repeated simulations and data collection was controlled using MATLAB.

2.3 Results

In order to determine how selection interference in small, finite populations operates, I decided to concentrate on two scenarios. First, I investigated how the initial frequency of B affects the fixation probability of A, and also how the final frequency of the most favourable genotype AB changes if the recombination rate r is increased.

Figure 2.1 plots how the fixation probability of A is affected by the initial frequency of the linked beneficial allele B, in a diploid population of size $2N = 1024$, with no recombination ($r = 0$) and the selective advantage of B is greater than that of A ($Ns_1 = 1$, $Ns_2 = 4$). It is clearly seen that the fixation probability of A is lowest if the initial frequency of B lies at an intermediate level. By fitting a cubic line of best fit to the data, calculated using least-squares regression, it was estimated that this minimum fixation rate occurred at $q_0 \approx 0.38$. The reason for this phenomenon is presumably that when the initial frequency of B is low, it is quickly lost stochastically, so Ab fixes as the fittest genotype. If the initial frequency of B is high then the fittest genotype AB dominates, so A also fixes through genetic hitchhiking. For intermediate values of B, however, there is conflict between the equally frequent Ab and aB genotypes. These genotypes then compete for fixation in the population, ultimately meaning that A is likely to be lost as aB fixes instead. This replacement of an advantageous allele with a fitter type was investigated analytically by BARTON (1995b).

Next, it is investigated how the fixation probability of the four genotypes AB, Ab, aB and ab changes as the level of recombination Nr increases. The initial frequencies of the two fitter alleles A and B were set at 0.1 each, and simulations were run for different values of s_1 and s_2. In the case of a small population ($2N = 64$), then an apparent advantage to recombination is observed when the selection strengths are relatively strong ($s_1 = 0.4$, $s_2 = 0.3$, equating to $Ns_1 = 12.8$, $Ns_2 = 9.6$). Figure 2.2 shows that as the recombination rate increases, the fixation probability of AB increases and that of Ab decreases. This is because recombination not only increase the fixation

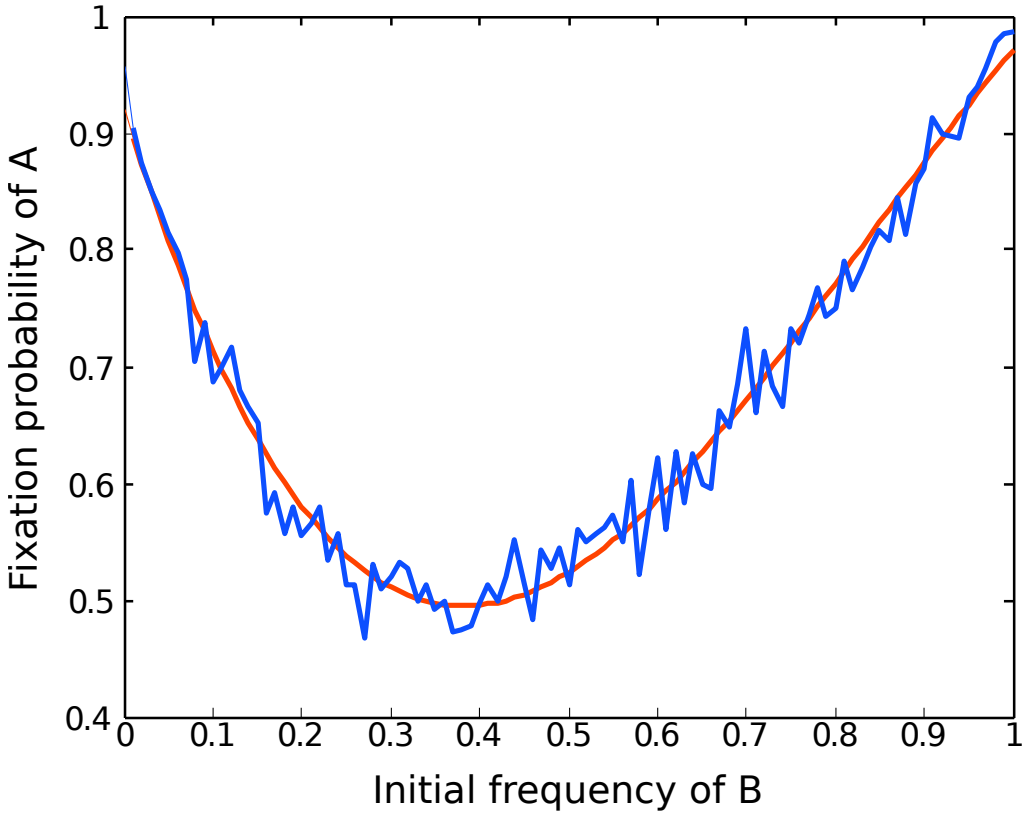


Figure 2.1: Plot of the fixation probability of A as a function of the initial frequency of B. Results from simulations are plotted (blue line), along with a least-squares fit to the data (red line). Parameters were $p_0 = 0.3$, $Ns_1 = 1$, $Ns_2 = 4$, $Nr = 0$ and $2N = 1024$. Note that due to the high initial frequency of A, its fixation probability is greatly higher than $2s_1$, which is the expected value if the allele arised with initial frequency $p_0 = 1/N$ (HALDANE 1927).

probability of B as outlined above, but is more likely to place the two beneficial alleles AB together, in line with the Fisher-Muller hypotheses (FISHER 1930; MULLER 1932). In theory, this should lead to increased levels of recombination evolving between the selected loci. However, in order to verify this, the simulation needs to be adjusted by adding a third locus that controls the rate of recombination between the two selected loci (as in OTTO and BARTON (1997)). Nevertheless, this result highlights a key feature of the Hill-Robertson effect: recombination breaks down negative linkage disequilibrium,

and hence the frequency of intermediate genotypes Ab and aB, increasing the frequency of the fittest genotype AB. In this particular example, A fixes in all cases as the fittest of the two beneficial alleles, but increased recombination aids the fixation of B.

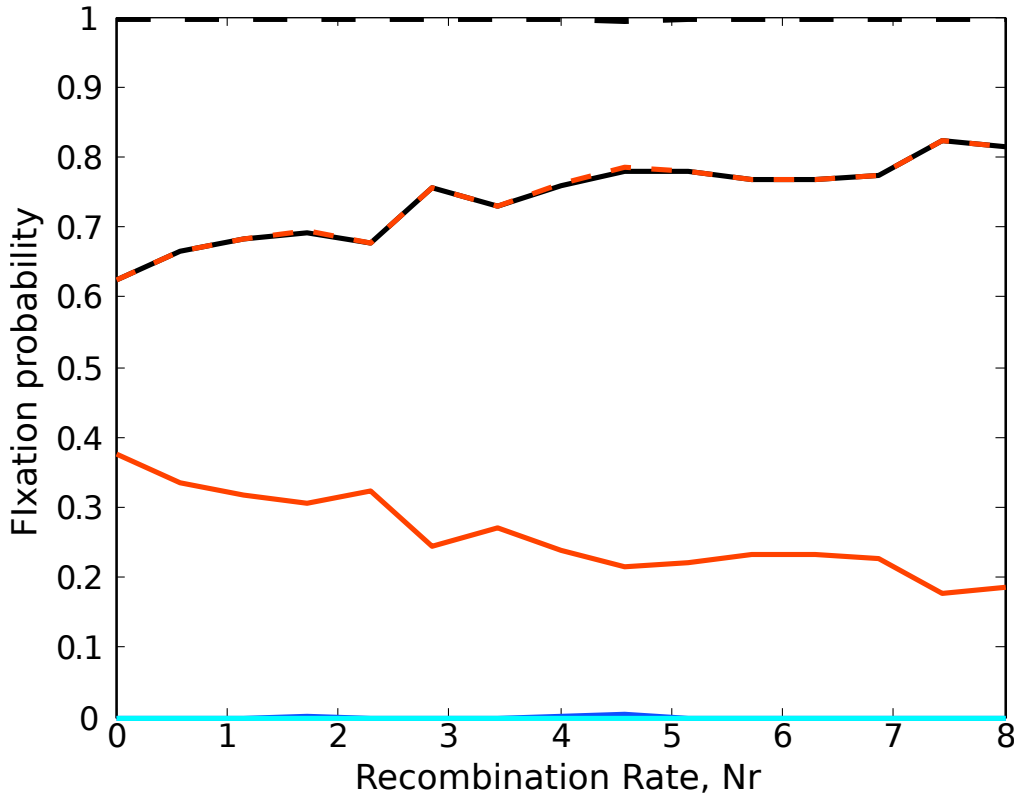


Figure 2.2: Plot of the fixation probability of each genotype; AB (black solid line); Ab (red solid line); aB (light blue line); and ab (dark blue line), as well as the favoured alleles A (black dashed line) and B (red dashed line). The parameters are $2N = 64$, $s_1 = 0.4$, $s_2 = 0.3$, $p_0 = q_0 = 0.1$.

However, an apparent disadvantage to recombination is observed in larger populations, where the selection coefficient of one allele is larger than the other. Figure 2.3 shows the results of a simulation for $2N = 128$, $Ns_1 = 6.0$ and $Ns_2 = 1.5$, again with $p_0 = q_0 = 0.1$. Whilst the fixation probability of A remains at a high value irrespective of the recombination rate, the fixation probability of AB *decreases* with increased recombination, and that of the intermediate genotype Ab increases. This behaviour arises

due to a higher initial frequency of AB genotypes, caused by both the high initial frequencies of the individual A and B alleles, and the additive fitness model assumed in this analysis. Overall, this leads to a decrease in the fixation probability of B. In this case, the formation of negative linkage disequilibrium through drift is not too strong, due to the stronger selection acting on A than B. Therefore, there is no major advantage to breaking down selection interference. Instead, recombination breaks apart the favourable genotype AB, forming an excess of Ab genotypes, which fix more often if recombination rates are high. In this scenario, it is expected that the creation of this 'recombination load' would disfavour a modifier for increased recombination, as also found by CHARLESWORTH *et al.* (1977).

2.4 Discussion

In this chapter, a simulation of selection acting on two linked loci was produced, which mimics the one outlined in HILL and ROBERTSON (1966). This is briefly analysed to show when selection interference is present due to low recombination rates, and how recombination can aid the formation of the fittest genotype. It also highlights a scenario when recombination would be disadvantageous; this arises if selection acting on one site is significantly higher than that on the second site. In this particular case, recombination breaks apart copies of the fittest genotype AB and increasing the fixation probability of less-fit genotypes instead.

This simulation is offered to show how selection and drift interfere with selection acting on linked loci, impeding the overall response to selection. This is an important process since breaking apart interference across multiple linked loci can produce strong selection for increased recombination (ILES *et al.* 2003; KEIGHTLEY and OTTO 2006; Chapter 3). The impact of selection interference between two linked sites was investigated analytically in BARTON (1995b), with OTTO and BARTON (1997); BAR-

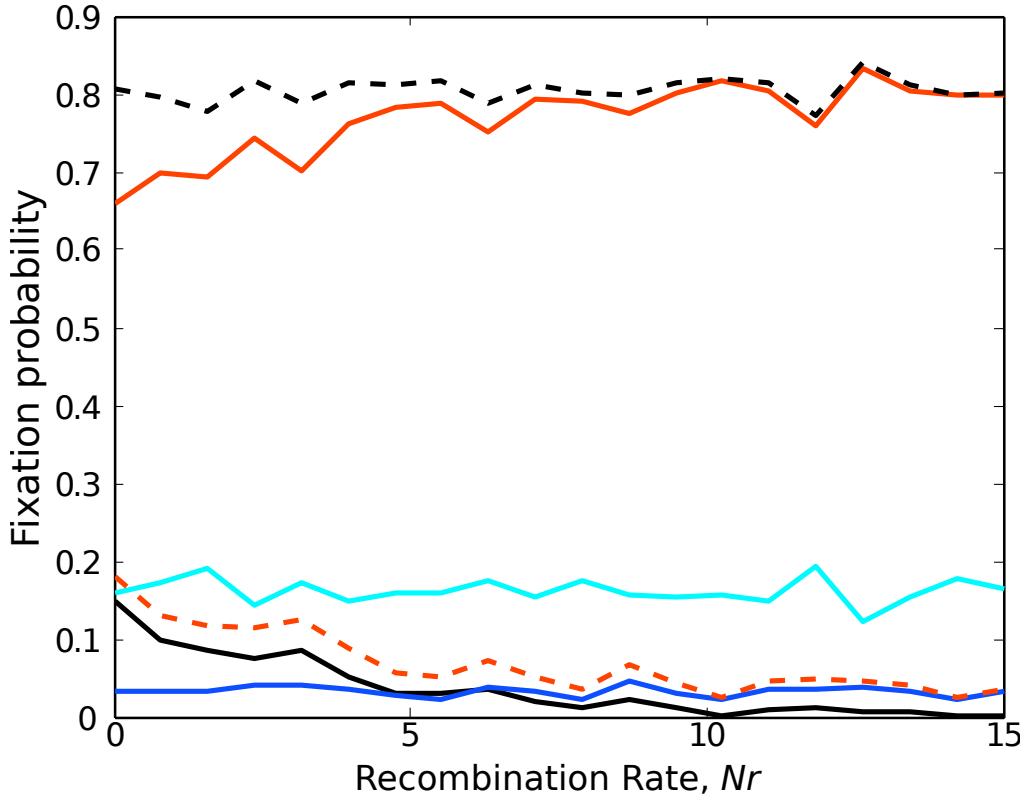


Figure 2.3: Plot of the fixation probability of each genotype; AB (black solid line); Ab (red solid line); aB (light blue line); and ab (dark blue line), as well as the favoured alleles A (black dashed line) and B (red dashed line). The parameters are $2N = 128$, $Ns_1 = 6$, $Ns_2 = 1.5$, $p_0 = q_0 = 0.1$.

TON and OTTO (2005) demonstrating how breaking apart selection interference leads to increased levels of recombination at a modifier locus.

2.A Derivation of selection equations

Here, I outline the equations used to model the change in genotype frequencies over time. These are based on well-known selection recursion equations, as used in HILL and ROBERTSON (1966); they are outlined here to demonstrate how these basic equations are derived.

Consider two loci, one with alleles A and a, the other with alleles B and b. The selection differences between these alleles are s_1 and s_2 respectively in homozygotes, and half that when present as a heterozygote. So, for example, a diploid individual with genotype AB/Ab has fitness $(1 - s_2/2)$, whilst the genotype AB/aB has fitness $(1 - s_1/2)$. It is assumed that selection is additive between loci. Table 2.1 outlines the frequency of all possible genotypes, along with their fitness and what gametes each one produces, according to a deterministic model.

Parental genotypes	Frequency	Fitness	Gametes produced			
			AB	Ab	aB	ab
AB/AB	f_1^2	1	1			
AB/Ab	$2f_1f_2$	1 - s_2/2	1/2	1/2		
AB/aB	$2f_1f_3$	1 - s_1/2	1/2		1/2	
AB/ab	$2f_1f_4$	1 - $s_1/2$ - $s_2/2$	1/2$(1-r)$	1/2 r	1/2 r	1/2$(1-r)$
Ab/Ab	f_2^2	1 - s_2		1		
Ab/aB	$2f_2f_3$	1 - $s_1/2$ - $s_2/2$	1/2 r	1/2 $(1-r)$	1/2 $(1-r)$	1/2 r
Ab/ab	$2f_2f_4$	1 - $s_1/2$ - s_2		1/2		
aB/aB	f_3^2	1 - s_1			1	
aB/ab	$2f_3f_4$	1 - s_1 - $s_2/2$			1/2	1/2
ab/ab	f_4^2	1 - s_1 - s_2				1

Table 2.1: Table of frequencies of gametes produced from parental genotypes under recombination c.

The change in gene frequencies under selection can then be computed from the values provided in Table 2.1. The new frequencies after selection and recombination can be found by summing the product of the genotype frequencies with the relative fitness of it (to represent the change in genotype frequency due to selection, where the relative fitness is the fitness of the genotype scaled by the mean value), and the proportion of gametes produced of the genotype of interest. So, for example, the change in frequency of AB under selection and recombination is calculated as:

$$
\begin{aligned}
\overline{w} f_1' &= f_1^2 + 1/2 \cdot 2 f_1 f_2 \cdot (1 - s_2/2) + 1/2 \cdot 2 f_1 f_3 \cdot (1 - s_1/2) \\
&\quad + 1/2 \cdot (1 - r) \cdot 2 f_1 f_4 \cdot (1 - s_1/2 - s_2/2) \\
&\quad + 1/2 \cdot r \cdot 2 f_2 f_3 \cdot (1 - s_1/2 - s_2/2) \\
&= f_1(f_1 + f_2 + f_3 + f_4) - (1/2) f_1((f_2 + f_4) s_2 + (f_3 + f_4) s_1) \\
&\quad - r(f_1 f_4 - f_2 f_3)(1 - s_1/2 - s_2/2) \\
&= f_1 - (1/2) f_1(s_1(1 - p) + s_2(1 - q)) - rD(1 - s_1/2 - s_2/2)
\end{aligned}
$$

$\overline{w}$ is the mean fitness of the population; this is the sum of the frequency of each genotype with the fitness of it:

$$
\begin{aligned}
\overline{w} &= \sum_{i,j} f_i f_j w_{ij} \\
&= f_1^2 + 2 f_1 f_2 (1 - s_2/2) + 2 f_1 f_3 (1 - s_1/2) \ldots
\end{aligned}
$$

where w_{ij} is the fitness of the genotype with frequency $f_i f_j$. Repeating the above calculation for all genotypes produces Equations 2.1 - 2.4 in the main text, which determine the change in genotype frequencies under selection and recombination.

Chapter 3

The role of advantageous mutations in enhancing the evolution of a recombination modifier

Published as: Hartfield, M., Otto, S.P. and Keightley, P.D., 2010. The Role of Advantageous Mutations in Enhancing the Evolution of a Recombination Modifier. *Genetics*, **184**(4): 1153–1164.

Chapter abstract

Although the evolution of recombination is still a major problem in evolutionary genetics, recent theoretical studies have shown that recombination can evolve by breaking down interference ('Hill-Robertson effects') among multiple loci. This leads to selection on a recombination modifier in a population subject to recurrent deleterious mutation. Here, I use computer simulations to investigate the evolution of a recombination modifier under three different scenarios of recurrent mutation in a finite population: (1) mutations are deleterious only, (2) mutations are advantageous only, and (3) there is a mixture of deleterious and advantageous mutations. I also investigate how linkage disequilibrium, the strength of selection acting on a modifier, and effective population size change under the different scenarios. I observe that adding even a small number of advantageous mutations increases the fixation rate of modifiers that increase recombination, especially if the effects of deleterious mutations are weak. However, the strength of selection on a modifier is less than the summed strengths had there been deleterious mutations only and advantageous mutations only.

3.1 Introduction

Sex and recombination between genomes is ubiquitous in nature, yet explaining their evolution has not proved to be easy (see recent reviews by HADANY and COMERON (2008) and OTTO (2009)). Recombination leads to the break-up of beneficial gene combinations (BARTON and CHARLESWORTH 1998), implying that offspring may suffer a recombination load (CHARLESWORTH and CHARLESWORTH 1975). Extra costs are incurred if sexual reproduction is also considered, such as the famous 'two-fold cost' (MAYNARD SMITH 1978); sexual offspring need two parents whereas asexuals have only one, so the latter can outgrow and outcompete sexuals.

Considering all of the associated costs, it has proved difficult to explain why recombination and sex are so common amongst eukaryotes. One hypothesis is that recombination breaks down 'Hill-Robertson' effects in asexuals (HILL and ROBERTSON 1966), which otherwise impede the response to selection. Hill-Robertson effects are the manifestation of many phenomena (discussed further in CHARLESWORTH *et al.* (2009)), including hitchhiking (MAYNARD SMITH and HAIGH 1974), background selection (CHARLESWORTH *et al.* 1993), and the accumulation of deleterious mutations by Muller's Ratchet (MULLER 1964; FELSENSTEIN 1974). Interference generates negative linkage disequilibrium (i.e., the accumulation of good alleles on bad genetic backgrounds), which reduces genetic variation in fitness compared to a population without this linkage disequilibrium (BARTON 2009). Interference also reduces the effective population size, N_e (ROBERTSON 1961; COMERON *et al.* 2008), because offspring from the same, fittest, lineages tend to be favoured. Recombination can increase the genetic variance in overall fitness, which can improve the response to selection (FISHER 1930; MAYNARD SMITH 1988b). If a modifier for increased recombination facilitates the production of fitter offspring in this way, then it has an indirect selective advantage and increases in frequency by virtue of being associated with fitter genotypes, in the spirit of Weismann's classic theory on the evolution of sex and recombination (WEISMANN

1887; BURT 2000).

Research into the Hill-Robertson effect has increased in recent years, with the development of analytical frameworks to study effects of drift at multiple loci. By extending earlier models that focused on selection alone (OTTO and FELDMAN 1997; BARTON 1995a), recent work has assessed how linkage disequilibrium, created by genetic drift and interference with selection, drives the evolution of a recombination modifier (BARTON and OTTO 2005). Indirect selection on a modifier also arises because recombination increases the probability that beneficial mutations establish within a population, and the strength of this selection has been modeled using multi-type branching processes (OTTO and BARTON 1997; ROZE and BARTON 2006).

Simultaneously, the ability to simulate large numbers of linked loci has increased, making it possible to evaluate the importance of the Hill-Robertson effect with selection acting across a genome. Simulations have demonstrated how breaking down interference can offer substantial selection on a modifier of recombination. OTTO and BARTON (2001), for example, showed that if a recombination modifier acts on loci experiencing directional selection, the effects of drift (which creates interference between loci) account for more selection on recombination than the effects of epistasis, in both 3-locus and 11-locus simulations. With only three loci, however, a modifier was not favoured in populations of $N > 10,000$ chromosomes without epistasis. Subsequently, ILES *et al.* (2003) demonstrated that as the number of loci under directional selection increases, strong selection for recombination occurs in even larger populations, including the largest population size considered ($100,000$ haploid individuals). Furthermore, with population structure, breaking down Hill-Robertson effects remains important for modifier evolution even in a population consisting of an infinitely-large number of finite-sized demes (MARTIN *et al.* 2006). Similarly, KEIGHTLEY and OTTO (2006) showed that a recombination modifier is strongly favoured in a population subject to recurrent deleterious mutation. The effects of drift again overwhelmed epistasis, and the dramatic

reduction in the effective population size at a neutral site, N_e, in asexual populations highlighted how Hill-Robertson effects impede natural selection (e.g. a reduction from $N = 50,000$ to $N_e \approx 100$ was observed in one example, if selection acting against deleterious mutants equalled 0.01 and there was complete linkage between loci). Selection on a sex modifier was also sufficiently strong that it could overcome a two-fold cost, but only if sex was initially rare and the modifier led to modest increases in the frequency of sex.

These advantages of recombination have been supported by experiments demonstrating that recombining populations are more responsive to selection. A study by MALMBERG (1977) found that allowing the bacteriophage T4 to exchange segments of its genome improved its rate of adaptation. More recent studies provide evidence that recombination increases the realised selection strength and fixation rate of new mutants in *Drosophila melanogaster* (RICE and CHIPPINDALE 2001); regions of the genome lacking in recombination in *Drosophila* show signs that selection has been impeded, whereas regions that have normal levels of recombination appear to be adapting more quickly (BETANCOURT *et al.* 2009; PRESGRAVES 2005; CHARLESWORTH *et al.* 2009); sex overcomes clonal interference in *Chlamydomonas reinhardtii*, which then accelerates adaptation (COLEGRAVE 2002); sex in stressed environments of yeast increases population variance in fitness and the response to selection (GODDARD *et al.* 2005); genetic drift induced by population bottlenecks in the RNA Bacteriophage $\Phi 6$ hampers the response to selection to a greater extent in asexuals then sexuals (POON and CHAO 2004); recombining populations of *Escherichia coli* are better able to break down interference between a known beneficial allele and other sites under selection, thereby increasing the rate of fixation of the fitter allele (COOPER 2007); *C. elegans* evolves outcrossing if a population is subject to an increased mutation rate or the presence of a pathogen, indicating that sex improves the response to adaptation (MORRAN *et al.* 2009).

Recently, a 'pluralist' framework has been proposed (WEST *et al.* 1999) arguing that multiple mechanisms work together to facilitate the widespread evolution of genetic recombination. A recent example investigating this theoretically was undertaken by OLIVEIRA *et al.* (2008), which aimed to gauge what degree of selection strengths acting on deleterious mutants drove the evolution of recombination via the process of background selection or through Muller's ratchet.

Here, I use computer simulations to extend the work of KEIGHTLEY and OTTO (2006) by considering both deleterious and advantageous mutations arising throughout the genome, in the spirit of a more 'pluralist' approach to the Hill-Robertson effect. Whereas KEIGHTLEY and OTTO (2006) demonstrated that genetic recombination is selectively favoured if multiple linked loci within the genome are subject to recurrent deleterious mutation, ILES *et al.* (2003) demonstrated that similar advantages to a modifier occur if multiple linked loci are subject to recurrent advantageous mutations. This motivates the question: in genomes subject to both deleterious and advantageous mutations, is recombination favoured more compared to a population subject to only one type of mutation and, if so, by how much? Specifically, I investigate whether the benefits of a recombination modifier in the presence of deleterious and advantageous mutations are additive. That is, if the modifier has selection advantage s_{Md} if just recurrent deleterious mutations occur at rate U_d, and advantage s_{Ma} if just advantageous mutations occur at rate U_a, then with both type of mutations occuring at a total rate of $U_d + U_a$, additivity implies a selective advantage of the modifier equal to $s_{Md} + s_{Ma}$. This is a reasonable null hypothesis to start with, and it allows me to test for the presence of interference between different types of mutation.

At sites under selection, the extent to which nucleotide substitutions are driven by positive selection or occur despite negative selection has been the topic of long debate. The proportion of advantageous mutations is likely to depend strongly on the match between the species and its current environment (for a recent review see EYRE-

WALKER (2006)). Comparing the Chimpanzee and human genomes suggests that hominids have experienced little adaptive evolution at the molecular level (THE CHIMPANZEE SEQUENCING AND ANALYSIS CONSORTIUM 2005; ZHANG and LI 2005) (although low rates of adaptive molecular evolution inferred in these species could be a consequence of population bottlenecks, which would downwardly bias estimates (EYRE-WALKER and KEIGHTLEY 2009)). On the other hand, BIERNE and EYRE-WALKER (2004) inferred that approximately 45% of amino acid substitutions in *Drosophila* are a consequence of adaptive evolution. This equates to one substitution in the genome every 45 years ($\sim$450 generations) (SMITH and EYRE-WALKER 2002), although these estimates are subject to discussion (SELLA *et al.* 2009). Here, I use rates of advantageous mutation based on these data to determine the role that adaptive mutations might play in the evolution of genetic recombination, especially when there are background deleterious mutations as well.

3.2 Methods

3.2.1 Simulation of a recombination modifier

The simulations start with a population of N mutant-free haploid chromosomes, each consisting of 100 equally spaced linked loci subject to recurrent mutation, unless stated otherwise. A new generation is created by selection, recombination (if present), and mutation to produce N offspring.

Three scenarios are investigated: mutants are exclusively deleterious (as in KEIGHTLEY and OTTO (2006)); mutants are exclusively advantageous (similar to ILES *et al.* (2003), although they considered standing variation only), or a proportion $x = k/s_a$ of mutants are advantageous and $1 - x$ are deleterious, with k being a numerical constant and s_a the selection coefficient for an advantageous mutant. The function $x = k/s_a$ reflects the assumption that strongly advantageous mutants are less likely to appear

than weakly selected ones (ANDOLFATTO 2007; JENSEN *et al.* 2008). The number of mutants is chosen from a Poisson distribution with mean U, except for the case where all spontaneous mutations are advantageous. Specifically, I assume that advantageous mutations are always a small proportion, x, of all mutations that occur. Thus, when only advantageous mutations are present, they occur at a rate Ux (this is equivalent to setting $s_d = 0$ in the case with both advantageous and deleterious mutants). If both advantageous and deleterious mutations are present, the overall mutation rate is not quite the sum of mutation rates from the separate scenarios; the overall mutation rate is U whereas the summed rate if mutants are deleterious only and advantageous only equals $U + Ux = (1 + x)U$. However x is assumed to be small (it is always less then $< 3\%$) and simulations with a total deleterious and advantageous mutation rate of $(1 + x)U$ give indistinguishable estimates of s_M, the selection strength on the modifier, compared to simulations with an overall mutation rate U (Figure 3.6 in supplementary figures).

In all scenarios, each site is equally likely to acquire a new mutation. Fitness effects of loci are multiplicative, with advantageous mutants having fixed fitness effects s_a and deleterious mutants have fitness effects of s_d. Thus with y advantageous mutants and z deleterious mutants, the fitness of a haploid individual equals $(1 + s_a)^y(1 - s_d)^z$. Fixed fitness effects are used to speed up simulations. Epistasis between mutations on the log-fitness scale is assumed to be absent, so that any increase in fixation rate of modifier mutations can be attributed to Hill-Robertson effects. Every 500 generations, the number of mutants is normalised; that is, the number of mutants present at a single locus is reduced by the minimum number of advantageous or deleterious mutants that any haploid individual possesses at that site in the entire population, so that the smallest number present at a single locus equals zero.

Except where noted, the population initially lacks genetic crossing over. To produce the next generation, a parent is chosen with replacement, with probability proportional to its fitness. This is then cloned and a number of mutants sampled from a Poisson

distribution is added to produce an offspring. This is repeated N times until the population is replenished. A burn-in of $5N$ such generations is run to allow the population variance to approach a steady-state. The state of the burn-in population is then saved, and a recombination modifier introduced at a randomly selected position on a randomly selected chromosome. The processes of selection, recombination and mutation are then repeated, except that two haploid parents are mated to allow crossovers to occur. The modifier increases the (Poisson) mean number of crossover events per chromosome during reproduction from $L = 0$ to $L = 0.1$ if it is present as a homozygote (and half that if it is heterozygous). The new modifier allele is then tracked until it is fixed or lost from the population. The process of introducing a single modifier mutation is repeated $5N$ times for each saved burn-in population and the total number of fixations is divided by $5N$ to obtain the fixation probability u. The statistic used to determine the selective advantage of the modifier is u/u^*, where $u^* = 1/N$, the fixation probability of a neutral mutation (KIMURA 1983). The above constituted one 'run' to produce a single statistic. Each run is executed 100 times from separate burn-ins to produce a distribution of fixation probabilities.

3.2.2 Parameter values used

The per-chromosome mutation rate (if mutants are solely deleterious, or deleterious and advantageous) is set to either $U = 0.1$ or 0.5, which are in the range of estimated deleterious mutation rates per chromosome in *Drosophila* (HALLIGAN and KEIGHTLEY 2006; HAAG-LIAUTARD *et al.* 2007; KEIGHTLEY *et al.* 2009). These values should also be similar to the joint deleterious and advantageous mutation rates, since selected mutants are believed to be mainly deleterious (CROW 1970). x, the proportion of mutants that are advantageous, is set to k/s_a with $k = 0.00023$. This value of k is chosen so that there was, on average, one substitution every 450 generations (a rate inferred for the *Drosophila* genome by BIERNE and EYRE-WALKER (2004)), in simulations that I conducted

with a small population ($N = 100$), low mutation rate ($U = 0.1$), with medium-strength advantageous and deleterious mutations both present ($s_a = s_d = 0.025$) and complete linkage between loci. This value of k is then used in all simulations investigated, however this will lead to higher rates of substitution occurring in simulations with large population sizes or mutation rates.

Values of s_a are set to 0.01, 0.025 or 0.05. I wanted to ensure that $Ns_a \geq 1$ for all $N \geq 100$, so that the fate of mutations is not determined by the action of drift alone, even if Hill-Robertson effects are absent (KIMURA 1983). These values are therefore somewhat higher then those obtained from analysis of amino-acid substitution data from *Drosophila*, although there is some overlap (see reviews by WRIGHT and ANDOLFATTO (2008); SELLA *et al.* (2009)).The appearance of strong adaptive mutations is best representative of advantageous mutants occurring at non-synonymous sites, where the substitution rate and selection strength is highest (ANDOLFATTO 2005).

I investigated a wide range of s_d values, from 0 to 0.05. Again precise values of s_d are hard to obtain from observations; smaller values of s_d investigated match up with estimates obtained from LOEWE and CHARLESWORTH (2006), however GARCÍA-DORADO *et al.* (1999) found a mean s_d of approximately 0.2. This high value may have resulted from simplifications used in the Bateman-Mukai inference method (LYNCH and WALSH 1998). One should be aware though that due to the strongly leptokurtic distributions of s_d found empirically, there is a great deal of variance around such estimates and many s_d values would be lower than those used in these simulations. However, note that effective population sizes N_e in *Drosophila* are larger than the populations simulated in this chapter, so $N_e s$ values of deleterious mutations used in these simulations will be of a weaker effect than those found in nature.

3.2.3 Measuring linkage disequilibrium for an asexual and recombining population

The log-fitness associated with a chromosome is additive in these simulations, so standard models of the expression of phenotypic quantitative traits can be used to measure the difference in fitness variance between the total additive and genetic variance obtained (BULMER 1976, 1980; KEIGHTLEY and HILL 1987). To measure linkage disequilibrium in an asexual population, the frequency of each individual mutant is tracked. A 'garbage collection' routine is executed every ten generations to clear memory; mutants that have either become fixed or lost are removed from the population, and a note kept of how many new mutant alleles have fixed. There is a burn-in of $5N$ generations, after which the mean linkage disequilibrium is measured over $5N$ generations; $LD = V_A - V_g$ for genetic variance V_A and genic variance V_g of the log-fitness (KEIGHTLEY and HILL 1987). LD is the contribution to the genic variance of log-fitness due to multi-locus linkage disequilibrium (BULMER 1980) and can be computed as above using the following terms for V_A and V_g:

$$V_A = (\sum_{i=1}^{N} w_i^2)/N - (\sum_{i=1}^{N} w_i)^2/N^2 \tag{3.1}$$

$$V_g = s_a^2(\sum_{j=1}^{m} \widetilde{y_j}(1-\widetilde{y_j})) + s_d^2(\sum_{j=1}^{m} \widetilde{z_j}(1-\widetilde{z_j})) \tag{3.2}$$

where w_i is the log-fitness of the ith chromosome in the population, given by $s_a y_i - s_d z_i$ (for y_i, z_i the number of advantageous and deleterious mutants respectively, in genome $i \in N$). $\widetilde{y_j}$, $\widetilde{z_j}$ is the number of genomes that a particular mutant appears in, divided by the total population size; that is, they are the frequencies of a segregating advantageous or deleterious mutant at locus j (with m segregating loci overall). Each locus has no more then two alleles segregating at any one time in this simulation.

For a population with recombination, a new mutant has a map position attributed to it drawn from a uniform $[0, 1]$ distribution. During reproduction, the position of a crossover is drawn from the same distribution. If one crossover is chosen, allelic states are exchanged at sites where the map position exceeds the recombination distance. If two crossovers occur, the states of loci are swapped where the mutant map position lies between the two crossover points. More than two crossovers are unlikely (the probability of more than two occurring is 0.00015, with $L = 0.1$), therefore only up to two exchanges are considered.

3.2.4 Measuring the strength of selection on a modifier

To measure selection on a modifier, a modifier allele is introduced at a frequency of 50% into a population after a burn-in. Introducing the modifier at an intermediate frequency prevents its immediate loss (or fixation), which would otherwise bias the long-term estimate of selection by forcing it to equal zero for all generations following its premature loss (or fixation). After its introduction, a modifier is tracked for 200 generations or until it is fixed or lost. At each generation following its introduction, the change in modifier frequency ΔM is noted, and selection on the modifier is estimated using the weak-selection equation $s_M = \Delta M/(p_M q_M)$ (BARTON 1995a). The value of s_M at only the 200^{th} generation is taken as the overall strength of selection acting on the modifier. This is repeated for $5N$ modifiers per burn-in, so a distribution of average selection strengths is developed. This is repeated for 100 burn-ins.

Measuring N_e for asexual populations

In order to estimate N_e, a neutral, linked locus is inserted into the genome at a random position, (i.e., the possibility of it being telomeric or centromeric is allowed). This locus affects a quantitative trait, which has an initial effect of zero. After a burn-in, the effect

of this locus is changed in each individual by adding Gaussian noise each generation with a mean of zero and variance V equal to one. It can be shown that the equilibrium variance should be VN_e for such a neutral trait (LYNCH and HILL 1986). The simulation is left to run with Gaussian-distributed mutations occurring every generation at the neutral locus for a further $5N$ generations in order to reach equilibrium, at which point the variance (and N_e) is measured. Average N_e values from independent burn-ins are calculated to form an overall distribution.

3.3 Results

3.3.1 Effects of advantageous mutations on a recombination modifier

I first investigate the dynamics of a recombination modifier in the presence of different types of mutations (deleterious only, advantageous only, both deleterious and advantageous). As observed by KEIGHTLEY and OTTO (2006), I found that the relative fixation probability of a modifier (u/u^*) rises as N increases for all cases investigated. Also, $u/u^* > 1$ for all simulations, indicating that a recombination modifier is always favoured. Full results for all scenarios investigated are provided in Appendix 3.A.

Although advantageous mutants arise in these simulations at a low frequency (the proportion of advantageous mutants is $x = k/s_a$, so for $s_a = 0.01$ only 2.3% of mutations are advantageous), their occurrence still causes a high fixation rate of the modifier, even in the absence of deleterious mutations. For example, with an advantageous mutation rate of $Ux = 0.0115$ and $s_a = 0.01$, the relative fixation probability $u/u^* = 3.58$ for $N = 1,000$, which is only slightly lower than that observed with deleterious mutations only and $U = 0.5$. The extent to which beneficial mutations select for recombination is even greater in larger populations (for example, $u/u^* = 47.4$ for $N = 10,000$), which

is greater than the corresponding value for the deleterious mutations case and $U = 0.5$. By increasing s_a to 0.05 but holding constant the net effect of mutations by decreasing the beneficial mutation rate to Ux = 0.0023 (as $x = k/s_a$), recombination is even more favourable (e.g. $u/u^* = 81.9$ for $N = 10,000$). So in the absence of deleterious mutations, recombination offers substantial benefits in aiding the fixation of recurrent advantageous mutants across multiple loci, especially in large populations.

Figure 3.1 compares relative rates of recombination modifier fixation for cases with both advantageous and deleterious mutations present (for $N = 25,000$ and $U = 0.1$). I observe that the presence of advantageous mutations alongside deleterious mutations leads to a higher fixation probability of a recombination modifier than if mutations are solely deleterious. The highest u/u^* of 220 occurs for the case of weakest selection against deleterious mutations, i.e. $s_d \approx 1/N$ and strong selection in favour of advantageous mutants, i.e. $s_a = 0.05$. For stronger s_d, increased purifying selection acting against deleterious mutants leads to the loss of a larger fraction of advantageous mutations, reducing the extent to which they can contribute to Hill-Robertson interference (CHARLESWORTH 1994; PECK 1994; JOHNSON and BARTON 2002). Even if there are no advantageous mutations, strong purifying selection means that individuals carry few deleterious mutations in their genome due to the loss of most genetic backgrounds, leading to a reduction in N_e, causing a recombination modifier to behave as more nearly neutral (e.g. u/u^* = 1.32 with $N = 25,000$, $s_d = 0.05$ and $U = 0.1$, no advantageous mutation). The interaction between s_M, the selection coefficient of a modifier gene, and N_e is discussed further in Section 3.3.2.

Such fixation probabilities may depend on the number of linked loci present (ILES *et al.* 2003). Consistent with this, I observe that the fixation probability rises as I increase the number of linked loci from 10 to 50 (Figure 3.7 in the supplementary figures). However, it appears that fixation probabilities reach a plateau as the number of linked loci approaches 100, indicating that these simulations capture the maximum impact of

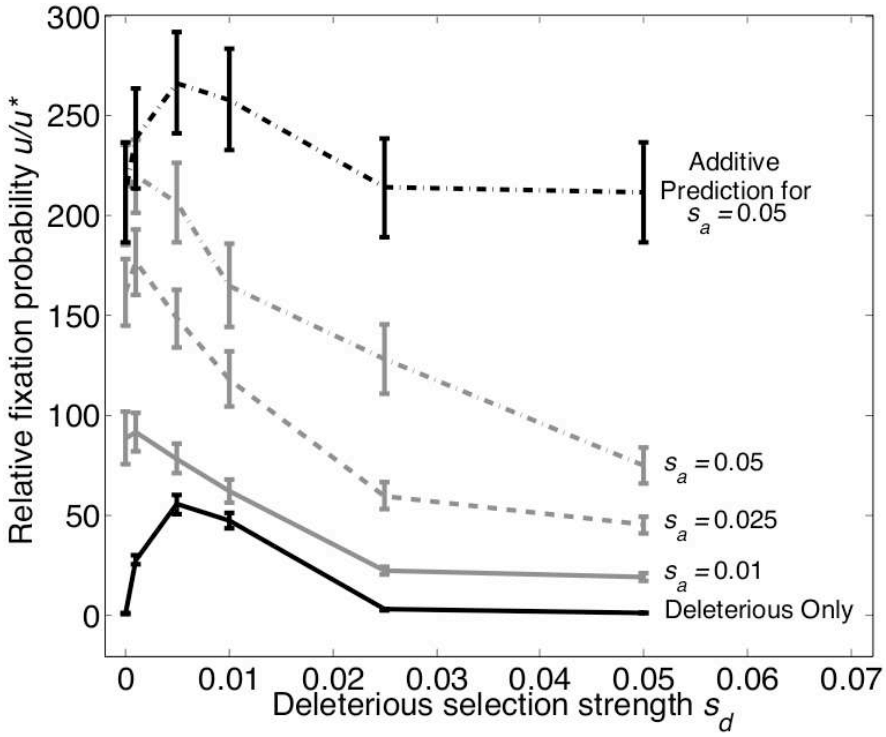


Figure 3.1: Relative fixation probability of a recombination modifier u/u^* for $N = 25,000$ as a function of the strength of selection acting against deleterious mutants. Mutations are just deleterious, or a mixture of deleterious and advantageous with strength $s_a = 0.01$, $s_a = 0.025$ or $s_a = 0.05$. These are compared to the expected u/u^* if both deleterious mutants and advantageous mutants ($s_a = 0.05$) are present and u/u^* is the sum of their independent fixation probabilities (with advantageous mutants only present, $u/u^* \sim$210 with $s_a = 0.05$). The chromosomal mutation rate in all cases is $U = 0.1$. Bars are 95% confidence intervals here and throughout the chapter.

Hill-Robertson effects in reducing the efficacy of selection, at least in the population sizes simulated.

With a combination of weak deleterious mutations and strong advantageous mutations, recombination offers a dual advantage, predominantly through the more efficient purging of deleterious mutants (stopping Muller's ratchet and reducing the mutation load (MULLER 1964; FELSENSTEIN 1974; KEIGHTLEY and OTTO 2006)), but also by aiding the fixation of rare advantageous mutants ('Fisher-Muller' hypothesis (FISHER 1930; MULLER 1932)). The increase in modifier fixation with higher s_a is likely to arise because strongly favoured mutants are likely to carry along with them many deleterious mutations in the absence of recombination (PECK 1994; HADANY and FELDMAN 2005) and recombination can free these advantageous mutations from their deleterious backgrounds. In line with this reasoning, Table 3.1 shows that for $N = 1,000$ and $U = 0.5$, recombination aids the fixation of advantageous mutants and decreases the fixation rate of deleterious mutants in all cases simulated.

3.3.2 Selection and fixation probabilities of a modifier

I next asked whether the benefits of the modifier brought about by purging deleterious mutants and fixing advantageous mutants are additive; recall that this means that if the modifier has selective advantage s_{Md} if just deleterious mutations occurs at rate U_d, and advantage s_{Ma} if just advantageous mutations occurs at rate U_a, then if both types of mutations occur at a total rate of $U_d + U_a$ the selective advantage is expected to equal $s_{Md}+s_{Ma}$. This is tested by comparing the selection coefficients at the 200th generation after the modifier is introduced for $N = 1000$, $U = 0.5$. Selection on the modifier is measured for three deleterious mutation selection strengths ($s_d = 0.01$, 0.025 and 0.05), where mutations are solely deleterious at rate U and again where both deleterious and advantageous mutations occur (with $s_a = 0.05$, $U = 0.5$). I also investigate the case where mutations are advantageous only (see points with s_d = 0; $s_a = 0.05$), which

Table 3.1: **Number of fixed mutants.** The average number of mutants that fix over $5N$ generations for $N = 1000$ and $U = 0.5$ given to three significant figures. Cases considered are those where mutations are solely deleterious, advantageous only, or both deleterious and advantageous. Fixations are measured for an asexual population or a population with a constant rate of recombination. Note that if advantageous and deleterious mutants are present, the strength relates to s_d, with $s_a = 0.05$. Figures in brackets are 95% confidence intervals here and throughout the chapter. When measuring the number of mutants fixed with recombination, the population recombines throughout the burn-in.

Case	Strength of deleterious mutations s_d		
	0.01	0.025	0.05
Only Deleterious Mutations			
Asexual Population	1260 (7.07)	823 (4.16)	471 (3.09)
Recombining Population	644 (5.99)	314 (3.57)	107 (1.71)
Only Advantageous Mutations			
Asexual Population	195 (2.17)	161 (1.77)	159 (1.43)
Recombining Population	378 (2.86)	362 (2.08)	371 (1.95)
Both Deleterious and Advantageous Mutations			
Number of deleterious fixed, asexual	1430 (8.88)	861 (5.09)	476 (3.14)
Number of deleterious fixed, recombining	985 (8.10)	393 (3.63)	125 (2.20)
Number of advantageous fixed, asexual	114 (1.91)	64.6 (1.54)	41.8 (1.27)
Number of advantageous fixed, recombining	269 (1.92)	149 (2.43)	90.1 (1.60)

occurs at the reduced rate $Ux = 0.0115$.

Results of this test are outlined in Figure 3.2. As with u/u^*, if the number of linked loci under selection is increased, s_M values appear to reach a plateau as the number of loci approaches 100 (Figure 3.8 in the supplementary figures). Whereas the addition of advantageous mutations enhances fixation of the modifier, the observed values of s_{Mb} (modifier strength when mutations are both advantageous and deleterious) fall short of the additive values, $s_{Md} + s_{Ma}$; in fact s_{Mb} is less than s_{Md} for all $s_d > 0.01$. s_{Mb} can exceed s_{Md} if the modifier is introduced at a low frequency ($< 10\%$) and $s_d \leq 0.1$ (Figure 3.9 in the supplementary figures), however s_{Mb} still falls short of the additive prediction. These selection coefficients are in contrast to the relevant fixation probabilities, u/u^*, as these values increase if advantageous and deleterious mutants are present, compared to the deleterious only case. However these fixation probabilities also act in a sub-additive manner (see also Appendix 3.A, and Figure 3.1). This decrease in s_M if advantageous and deleterious mutants are both present seem to verify the hypothesis that extra interference is present if two types of mutations are present together; breaking this down offers an increase. Selection on the modifier also increases with N (Figure 3.10 in the supplementary figures), consistent with the hypothesis that a recombination modifier is more strongly selected for in larger populations.

Interestingly, increasing s_d increases selection on the modifier, s_M, with or without advantageous mutations, whereas the fixation probability of the modifier u/u^* decreases with $s_d \geq 0.01$ in all simulations (Figure 3.1). The explanation of this paradoxical result is connected with changes in the effective population sizes N_e and how Hill-Robertson interference affects fixation of the modifier. The fixation probability of a new mutant is determined by its selection strength s and the effective population size N_e according to $u = (1 - exp(-2sN_e/N))/(1 - exp(-2sN_e))$ (KIMURA 1983). By reducing N_e, Hill-Robertson effects can reduce the fixation probability of a new mutation (the recombination modifier in this case), even if it is more strongly favoured (Table 3.2).

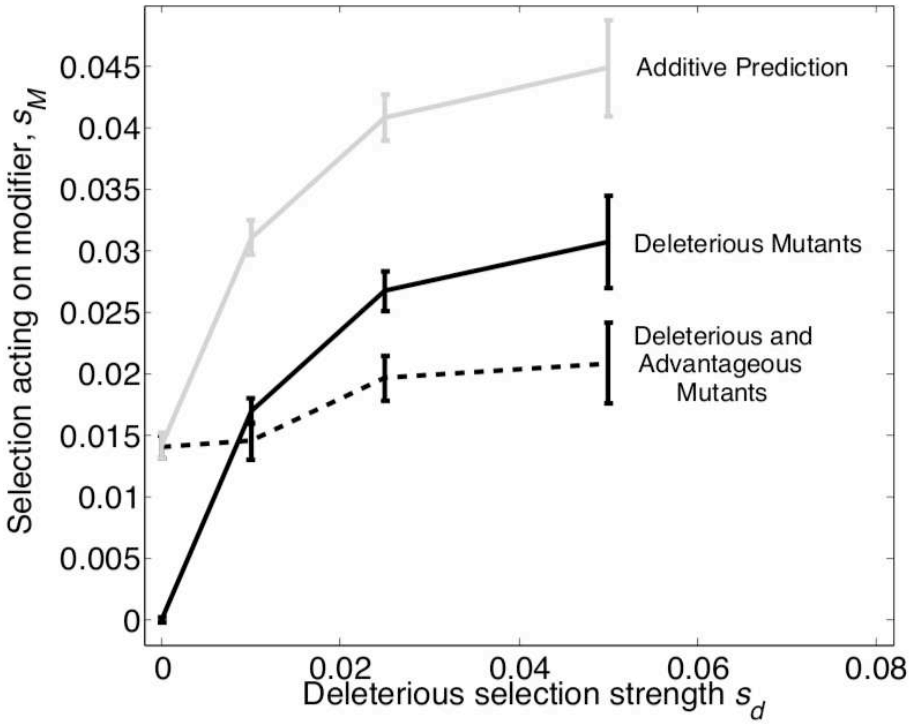


Figure 3.2: Selective advantage of a modifier, s_M, inferred for cases where mutations are deleterious only and where mutations are deleterious and advantageous. This is compared to the 'additive' prediction for the deleterious and advantageous case ($s_M \approx 0.013$ if mutations are solely advantageous, which is the result for the deleterious and advantageous mutant case if $s_d = 0$). $N = 1000$, $U = 0.5$, $s_a = 0.05$ if present.

With respect to a modifier, having more strongly selected deleterious mutations has a more dramatic impact on reducing N_e than increasing s_M, with the net result that the modifier is less likely to fix.

Table 3.2: **Estimates of s_M and N_e.** s_M and N_e are measured for different scenarios investigated ($N = 1000$, $U = 0.5$, s_M plotted in Figure 3.2), along with predicted fixation rates based on these values using Kimura's formula ('Pred.'). These are compared to fixation rate of the modifier u/u^* obtained from simulations.

Deleterious only case				
s_d	Modifier s_M	N_e	Pred. u/u^*	Observed u/u^*
0.01	0.0170 (0.0010)	117.88 (8.41)	4.073	4.518
0.025	0.0268 (0.0016)	73.08 (4.36)	3.989	4.000
0.05	0.0308 (0.0034)	55.14 (2.95)	3.508	3.660
Deleterious and advantageous mutants case				
s_d	Modifier s_M	N_e	Pred. u/u^*	Observed u/u^*
0.01	0.0146 (0.0015)	97.62 (5.26)	3.021	5.206
0.025	0.0197 (0.0018)	72.48 (3.80)	3.026	4.456
0.05	0.0209 (0.0033)	52.29 (2.60)	2.460	3.900

Whereas Kimura's formula offers accurate estimates of u/u^* if mutations are deleterious, it underestimates fixation probabilities if advantageous mutants are present as well. It appears that selective sweeps alter N_e as mutations rise in frequency, by reducing fitness variance at linked sites (MAYNARD SMITH and HAIGH 1974), violating the assumption that N_e is constant at steady state. This result also suggests that the presence of Hill-Robertson effects can increase the fixation probability of a modifier, relative to that expected at a single locus. This could be due to recombination increasing fitness variance as a modifier rises in frequency (see below), increasing N_e from the value in an asexual population as interference is broken down.

3.3.3 Testing the effectiveness of the diffusion approximation in predicting modifier fixation rates

The previous stochastic simulations are limited in the sense that they can only be run for population sizes that are small, compared to some of the large effective population sizes found in nature. In order to predict outcomes for larger N, I now investigate diffusion approximations. These predict that the behaviour of a new mutant is left unchanged if $N_e\mu$, N_es and N_er are constant and small (note that I refer to μ, the per site mutation rate, as opposed to U, the per chromosome mutation rate; $\mu = U/100$ and r is the recombination fraction between individual loci). The diffusion approximation should also hold if N_er is large and $N_e\mu$ and N_es are kept small but I do not focus on that situation here. A thorough overview of such work can be found in EWENS (2004) with an example provided by GORDO and CHARLESWORTH (2000). However, the diffusion approximation may not hold for this simulation, since N_e changes with different mutation rates and increases with higher rates of recombination.

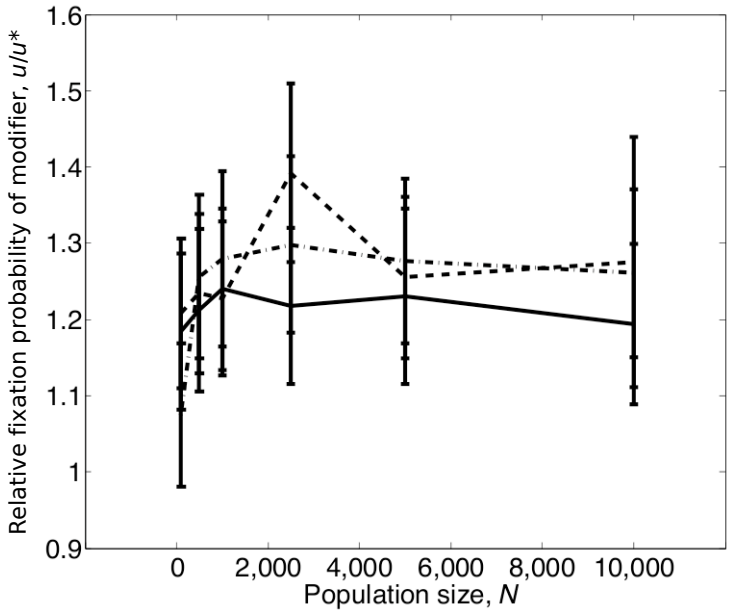
I decided to calculate u/u^* for $N\mu = 1$, $Ns = 5$ and $NL = 5$ for all cases of mutation (deleterious only; advantageous only; both deleterious and advantageous), in order to determine whether fixation is constant as a function of N. These results are plotted in Figure 3.3(a); the graph shows that u/u^* becomes approximately constant as a function of N, albeit at low values (around 1.25). This suggests that diffusion approximations might be useful as a guide to predict modifier behaviour for larger N than is possible to simulate directly. However, as Figure 3.3(b) shows, u/u^* increases non-linearly with N if simulations are run for large $N\mu = 10$, $NL = 1000$ and $Ns = 100$. These parameter values are chosen so that $U = 0.1$, $L = 0.1$ and $s = 0.01$ if $N = 10,000$. The non-linear estimates of u/u^* obtained implies that one cannot extrapolate simulation results unless N is much larger if rates of mutation, recombination and strength of selection are of these magnitudes (i.e. comparable to parameters observed for *Drosophila*). The observation that selection on the modifier is not invariant when $N\mu$, NL and Ns are held

constant but large could either be due to a breakdown in the diffusion approximations or due to changes in N_e caused by recombination reducing Hill-Robertson interference. These results may also reflect an instability in N_e due to an accumulation of deleterious mutations, as caused by Muller's ratchet (GESSLER 1995; GORDO and CAMPOS 2008).

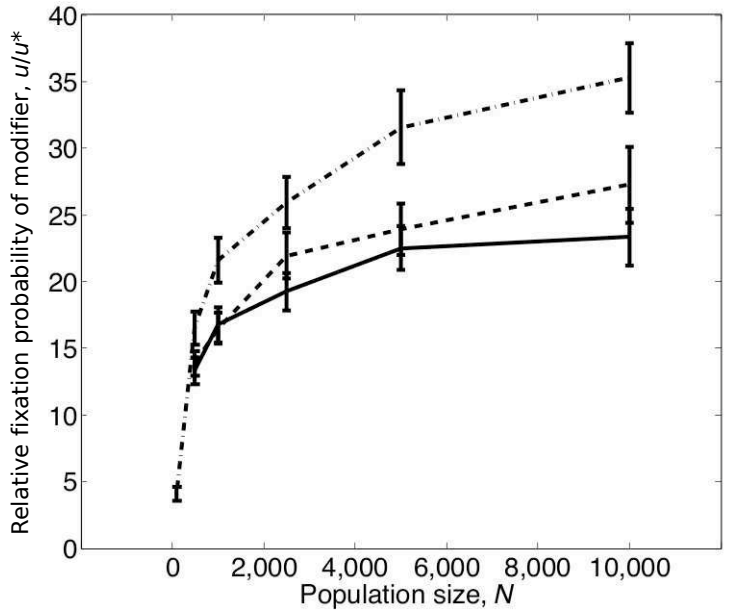
3.3.4 Effects of variance and linkage disequilibrium on modifier selection

In this section, I investigate how linkage disequilibrium changes with recombination and whether these values relate to u/u^*. Figure 3.4 compares the genic variance, genetic variance and variance due to linkage disequilibrium in asexual and recombining populations for two mutational cases (deleterious only; deleterious and advantageous) for $N = 5,000$ and $U = 0.1$. In both cases plotted, the genic and genetic variance is unchanged or it increases in the presence of recombination. With deleterious mutations only, the genetic variance is approximately $s_d^2(U/s_d) = Us_d$, the interference-free value of expected variance in an infinite population. If both advantageous and deleterious mutants segregate, the variance increases more substantially with recombination compared to the deleterious only case. By Fisher's fundamental theorem of natural selection, this increase in genetic variance should hasten the response to selection and improve the population mean fitness (FISHER 1930; PRICE 1972). Recombination is selected for through association with this rise in fitness.

In both panels the magnitude of linkage disequilibrium is only slightly different in a recombining population. For $s_d = 0.01$ to 0.025, where modifier fixation is greatest, the magnitude of linkage disequilibrium increases by approximately 10-fold if advantageous mutants are present alongside deleterious mutants, compared to the deleterious only case. This signifies a large amount of extra interference being created with the presence of advantageous mutations. However there is not a one-to-one correspondence between increases in linkage disequilibrium and increases in u/u^*. For example, with $s_d = 0.05$,

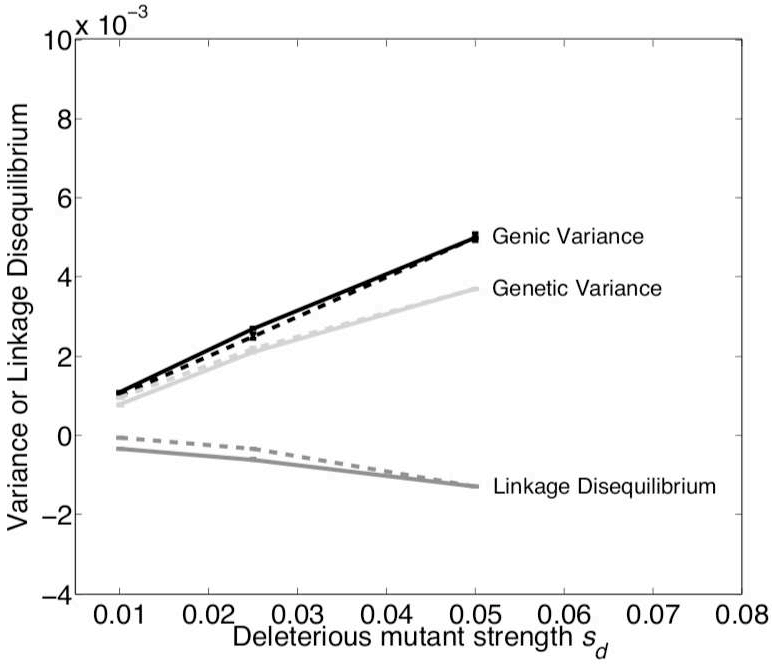


(a)

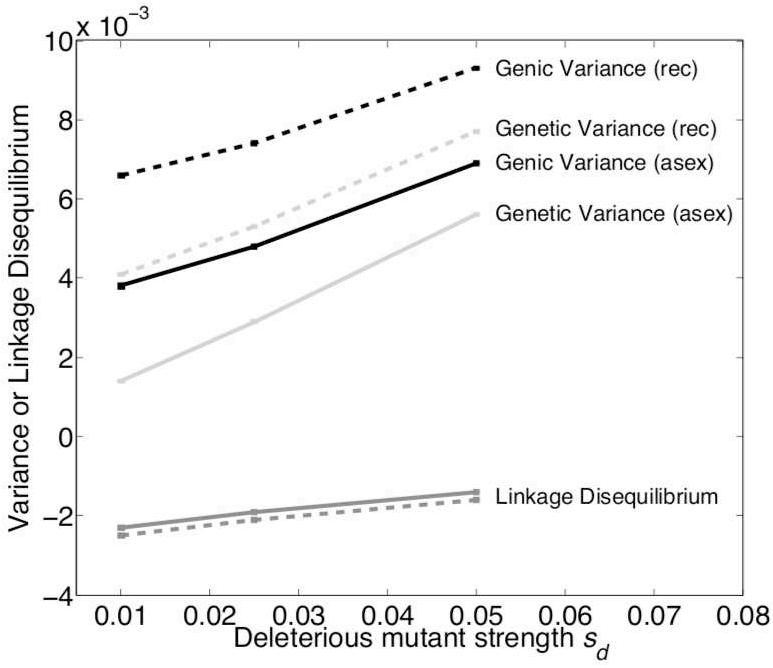


(b)

Figure 3.3: (a) u/u^* as a function of N for fixed $N\mu$, Ns, NL which are of $O(1)$ (μ is the per site mutation rate, $\mu = U/100$) if mutations are deleterious (solid line), advantageous (dot-dashed line) or deleterious and advantageous (dashed line). (b) u/u^* as a function of N with $N\mu$, Ns, NL fixed but no longer $O(1)$.



(a)



(b)

Figure 3.4: Genic variance, genetic variance and linkage disequilibrium for an asexual (solid lines) and recombining (dashed lines) population. (a) is for the deleterious only case and (b) for advantageous and deleterious mutations case (with $s_a = 0.05$). $N = 5000$, $U = 0.1$.

u/u^* is 11-fold higher in the presence of both advantageous and deleterious mutants than in the case with only deleterious mutants, despite there being little difference in the linkage disequilibrium present.

A possible explanation for this mismatch between the observed level of linkage disequilibrium and the fate of a modifier of recombination is that as recombination breaks down linkage disequilibrium, more advantageous alleles are rescued from poor genetic backgrounds, which increases their chance of establishment and creates extra interference. Due to this, the better predictor for the increase in modifier fixation rate is the increase in genic variance V_g within a population (BARTON and OTTO 2005). For example, with $s_d = 0.01$ genic variance increases by 0.0028 in a recombining population compared to an asexual population. If $s_d = 0.05$, the increase is only by a value of 0.0024. However the corresponding fixation probability u/u^* drops from 29.85 if $s_d = 0.01$ to 14.52 with $s_d = 0.05$. So, although the difference in genic variance between an asexual and recombining population decreases with stronger selection acting against deleterious mutants, the drop is not large enough to predict the steep decline in fixation probability associated with these parameter values.

3.3.5 Does the advantage of recombination continue to rise with U?

Interestingly, the relative fixation probability of a modifier does not rise linearly with the mutation rate (Figure 3.5(a)). This result holds even if the number of linked loci under selection is reduced (Figure 3.11(a) in the supplementary figures). This is unexpected, as one might assume that the fixation probability of a recombination modifier increases with higher U, since more mutants are produced that creates extra interference between sites. One reason for this behaviour is that as the mutation rate increases, the extent of Hill-Robertson interference also increases, reducing N_e and the spread of a modifier. Thus, even though selection on the modifier, s_M, rises with U (see Table 3.3 for the deleterious only case), the two effects cancel, leaving the fixation rate of the modifier

relatively constant as U increases. s_M values are approximately equal to those shown if there are fewer linked loci under selection (Figure 3.11(b) in the supplementary figures).

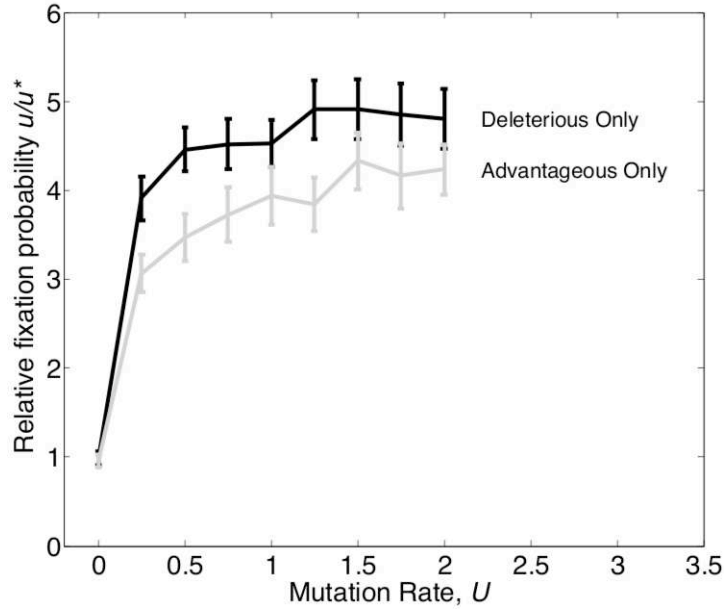
Table 3.3: **Estimates of s_M, N_e based on values at a neutral locus and predicted/observed u/u^* values as a function of U.**

$N = 1000$, $s_d = 0.01$				
U	Modifier s_M	N_e	Pred. u/u^*	Observed u/u^*
0.25	0.0119 (0.00067)	145.79 (4.51)	3.575	3.91
0.50	0.0172 (0.00097)	114.78 (3.28)	4.018	4.46
0.75	0.0213 (0.0010)	103.54 (3.11)	4.455	4.52
1.00	0.0220 (0.0013)	94.83 (2.77)	4.229	4.53
1.25	0.0230 (0.0016)	88.05 (2.61)	4.114	4.91
1.50	0.0240 (0.0017)	83.23 (2.34)	4.062	4.91
1.75	0.0250 (0.0017)	78.77 (2.03)	4.009	4.85
2.00	0.0254 (0.0018)	75.48 (2.12)	3.912	4.80

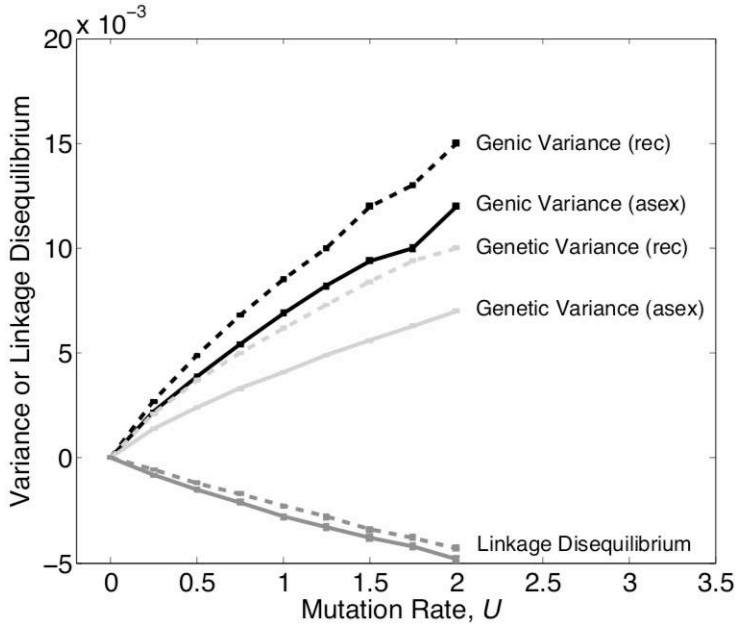
This argument is supported by investigating the underlying genetic and genic variances (Figure 3.5(b)). As U increases, the rise in the magnitude of genic variance with recombination becomes larger, indicating greater selection for the modifier. However the magnitude of linkage disequilibrium also increases due to the presence of more segregating polymorphisms, which will drive down the effective population size. Thus, while one might expect Hill-Robertson effects to select for recombination in direct proportion to the mutation rate, genetic interference is, to a large extent, self-limiting, and a strong diminishing returns relationship is observed, which tapers off once chromosome-wide mutation rates reach $\sim U = 1$.

3.4 Discussion

In this chapter, I show that for mutation rates and mean selection strengths that are representative of what is known in *Drosophila*, the presence of advantageous mutations



(a)



(b)

Figure 3.5: (a) u/u^* as a function of U if mutation is deleterious only and if mutation is advantageous only at rate Ux. $N = 1000$, $s_d = s_a = 0.01$. (b) Genic variance, genetic variance and linkage disequilibrium as a function of U where mutations are deleterious only, for the same parameters. Populations are asexual (solid lines) or recombining (dashed lines).

can lead to substantial selection on a modifier for recombination. As Figure 3.1 demonstrates, the highest advantages occur if s_d is low and s_a is high. Hence the low rate of adaptive amino-acid substitutions observed in *Drosophila* is capable of aiding the evolution of recombination and can help to account for its widespread occurrence.

That said, the addition of advantageous mutations alongside deleterious mutations increases the fixation of a modifier in a sub-additive fashion (Appendix 3.A and Figure 3.1). This demonstrates that whereas there can be a pluralist advantage for recombination in fixing beneficial alleles as well as purging deleterious mutants, the benefit gained from aiding selective sweeps is not as great as one might have expected if assuming that the modifier acts on deleterious mutants and advantageous mutants independently. This arises due to extra interference being created by deleterious mutants causing the loss of beneficial mutants, as highlighted in Table 3.1.

This chapter also offers insight into how to best measure the rate of evolution of a recombination modifier in the presence of Hill-Robertson interference. Figure 3.2 shows that a recombination modifier can be strongly selected for if introduced at 50% frequency; however, due to the high levels of Hill-Robertson interference present (Table 3.2), it only has a fixation rate that is slightly higher then a neutral mutant ($u/u^* = 1.32$). This suggests that measuring the strength of selection acting on a strong modifier can be misleading, since it does not take into account how Hill-Robertson interference impedes the spread of a beneficial mutant. This interference is broken down as a modifier increases in frequency in an initially asexual population, increasing N_e over time. However if there was already some recombination present in the population, or if the modifier is weak, then N_e would not appreciably change, so s_M might offer accurate insight into the fate of a recombination modifier in these cases.

I also demonstrate that, whereas it is theoretically possible to extrapolate fixation values of the modifier for larger N from fixation rates for small N using diffusion models, these assumptions will hold best if the values Ns, $N\mu$ and Nr are of $O(1)$, which

predict small rates of modifier fixation. Using larger N and parameter values, diffusion approximations breaks down, and thus I have to resort to full simulation.

By examining the genic and genetic variance in the simulations (Figure 3.4), I observe that negative linkage disequilibrium is created, which is indicative of Hill-Robertson interference (HILL and ROBERTSON 1966). Recombination increases genetic variance in fitness within a population, in line with existing theory on the evolution of a recombination modifier in the presence of drift (BARTON and OTTO 2005). These results, however, highlight an important point that even though linkage disequilibrium is indicative of interference, the magnitude of it does not determine the change in frequency of a modifier (BARTON 1995a; BARTON and OTTO 2005). This is exemplified when $s_d = 0.05$, where linkage disequilibrium values are similar in the presence and absence of beneficial mutants, yet modifiers of recombination are more strongly favoured in the latter case. This is because a particular level of genome-wide linkage disequilibrium (measured by $V_A - V_g$) can arise either when there are many segregating deleterious mutations (in which case, advantageous mutations are unlikely to fix), or when there are few segregating deleterious mutations and more advantageous mutations are able to establish.

Overall, this chapter demonstrates how beneficial mutations provide strong selection for a recombination modifier. However, there are a few caveats associated with the parameters used in this chapter, which should be investigated further in order to determine the full extent of the evolution of a recombination modifier.

Values of s_a used in these simulations are higher then those inferred for amino-acid substitutions in *Drosophila*, in order to prevent drift overwhelming mutations for small population sizes that I investigate. Using smaller values of s_a will certainly reduce the fixation probability of the modifier, at least for the population sizes investigated in these simulations.

Dominance of diploid deleterious mutations is also not considered here. ROZE

(2009) showed how, if deleterious mutants are highly recessive, recombination is selected against, because breaking apart multi-locus heterozygosity incur a fitness disadvantages, especially if selection on the mutations is weak. Future work should investigate how the presence of advantageous mutations affects this result, although such a study is likely to strongly rely on the dominance of beneficial mutations, which is only poorly known.

The strength of selection on adaptive mutations also depends on the organism under observation and the state of its environment. Relative fitness differences in *Drosophila* can be reduced if their populations are dense or if there is a lack of available food (KONDRASHOV and HOULE 1994). On the other hand, in bacteria and viruses, advantageous mutants with larger fitness effects have been observed in stressed environments (on the order of $s = 5$ (BARRETT *et al.* 2006) or even $s = 12$ (BULL *et al.* 2000)). Bearing all this in mind, the effects on a modifier over a larger range of selection parameters should be investigated.

These results also offer predictions as to when recombination can evolve. If an organism moves to a new environment to which it is maladapted, this model predicts that higher rates of recombination are more likely to arise in this new environment. Furthermore, if background deleterious mutations are frequent, recombination has an extra advantage in aiding purifying selection and is even more likely to evolve than in the presence of beneficial mutations alone. Such a scenario was discussed by HADANY and FELDMAN (2005) and could explain why recombination is more likely to occur in new, stressed environments (GRISHKAN *et al.* 2003; ABDULLAH and BORTS 2001).

Finally, I did not investigate whether the advantages to a recombination modifier in the presence of advantageous mutants transfer over to a sex modifier. This requires an adjusted model to account for the costs of sex (MAYNARD SMITH 1978) and to ensure that excessive inbreeding is avoided (if rare sexuals can only mate with other sexuals). This is a well-known problem with regards to the evolution of sex (see, for example,

PECK (1993)) and I will investigate this in Chapter 5.

3.A Full list of modifier fixation results

The following section outlines all results collected for the fixation probability of a recombination modifier, u/u^*. Data was collected for N = 100, 500, 1000, 5000, 10,000 and 25,000; $s_a = s_d$ = 0.01, 0.025 and 0.05; and U = 0.1 and 0.5. Data was also collected with s_d = 0, 0.001 and 0.005 if $N \geq 10,000$ and $U = 0.1$.

There are three cases; mutations are deleterious only, advantageous only, or deleterious and advantageous. Results will be presented in this order. Bracketed values are 95% confidence intervals here and throughout the appendix.

3.A.1 u/u^* for deleterious mutations only present

u/u^* as N increases, $U = 0.1$			
N	Deleterious mutations strength s_d		
	0.01	0.025	0.05
100	1.084 (0.087)	1.186 (0.099)	1.158 (0.090)
500	2.068 (0.147)	2.220 (0.182)	1.570 (0.141)
1000	3.147 (0.314)	2.912 (0.258)	1.550 (0.182)
5000	11.79 (0.797)	6.852 (0.609)	1.274 (0.109)
10000	23.34 (2.117)	6.646 (0.901)	1.260 (0.095)
25000	47.55 (3.639)	3.333 (0.302)	1.315 (0.103)
N	Deleterious mutations strength s_d		
	0.000	0.001	0.005
10000	1.060 (0.097)	9.988 (0.645)	22.30 (1.652)
25000	1.026 (0.146)	28.00 (2.226)	55.72 (4.887)
$U = 0.5$			
100	1.174 (0.096)	1.280 (0.100)	1.392 (0.116)
500	2.550 (0.186)	2.502 (0.225)	2.686 (0.220)
1000	4.518 (0.379)	4.000 (0.342)	3.664 (0.312)
5000	16.78 (1.737)	11.18 (2.095)	9.628 (1.102)
10000	31.73 (3.165)	23.59 (2.781)	13.76 (1.242)
25000	70.46 (7.854)	47.16 (4.990)	23.11 (2.620)

3.A.2 u/u^* for advantageous mutations only present

Note that in this case, the actual mutation rate is dependent on the selection strength s_a of the mutant, using the function $Ux = Uk/s_a$ as I am interested in the modifier fixation probability if advantageous mutants occur at the rate they would occur if arising alongside deleterious mutations. This enables us to test whether the selection strength of the modifier is additive compared to the advantageous only and deleterious only case.

u/u^* as N increases, mutation rate = $Uk/s_a = 0.1 \cdot 0.00023/s_a$			
N	Advantageous mutations strength s_a		
	0.01	0.025	0.05
100	1.030 (0.090)	1.074 (0.094)	1.072 (0.088)
500	1.390 (0.121)	1.866 (0.152)	2.274 (0.230)
1000	2.338 (0.173)	3.742 (0.326)	4.112 (0.414)
5000	16.33 (1.345)	25.36 (2.143)	34.11 (3.502)
10000	35.29 (2.625)	58.98 (5.651)	77.56 (7.671)
25000	88.88 (13.20)	161.77 (16.48)	210.74 (24.98)
Mutation rate = $Uk/s_a = 0.5 \cdot 0.00023/s_a$			
N	Advantageous mutations strength s_a		
	0.01	0.025	0.05
100	1.022 (0.075)	1.066 (0.088)	1.122 (0.103)
500	1.870 (0.121)	2.706 (0.225)	3.000 (0.270)
1000	3.576 (0.236)	4.598 (0.425)	6.442 (0.578)
5000	20.84 (1.780)	29.14 (2.623)	35.50 (4.118)
10000	47.44 (4.000)	68.76 (6.968)	81.92 (8.656)
25000	115.62 (12.09)	152.87 (15.99)	184.01 (18.30)

3.A.3 u/u^* for advantageous and deleterious mutations present

u/u^* for N = 100, $U = 0.1$			
s_a	Deleterious mutations strength s_d		
	0.01	0.025	0.05
0.01	1.076 (0.084)	1.124 (0.089)	1.188 (0.083)
0.025	1.060 (0.096)	1.114 (0.093)	1.164 (0.095)
0.05	1.092 (0.090)	1.198 (0.091)	1.240 (0.100)
$U = 0.5$			
0.01	1.212 (0.098)	1.226 (0.098)	1.238 (0.112)
0.025	1.164 (0.090)	1.342 (0.120)	1.260 (0.108)
0.05	1.202 (0.108)	1.206 (0.108)	1.330 (0.093)
u/u^* for N = 500, $U = 0.1$			
s_a	Deleterious mutations strength s_d		
	0.01	0.025	0.05
0.01	1.882 (0.148)	1.952 (0.164)	1.530 (0.143)
0.025	2.166 (0.174)	2.192 (0.169)	1.772 (0.160)
0.05	2.538 (0.228)	2.214 (0.177)	1.952 (0.176)
$U = 0.5$			
0.01	2.510 (0.190)	2.744 (0.209)	2.716 (0.216)
0.025	2.758 (0.218)	2.824 (0.206)	2.750 (0.242)
0.05	3.222 (0.250)	2.698 (0.220)	2.620 (0.222)
u/u^* for N = 1000, $U = 0.1$			
s_a	Deleterious mutations strength s_d		
	0.01	0.025	0.05
0.01	3.326 (0.236)	2.902 (0.203)	1.772 (0.171)
0.025	3.642 (0.244)	3.510 (0.294)	2.128 (0.178)
0.05	4.706 (0.377)	3.666 (0.344)	2.770 (0.254)

$U = 0.5$			
0.01	4.656 (0.419)	4.718 (0.374)	4.120 (0.430)
0.025	4.596 (0.362)	4.318 (0.385)	4.028 (0.322)
0.05	5.206 (0.445)	4.456 (0.436)	3.900 (0.321)

u/u^* for N = 5000, $U = 0.1$			
s_a	Deleterious mutations strength s_d		
	0.01	0.025	0.05
0.01	12.41 (1.139)	7.934 (0.708)	4.010 (0.253)
0.025	19.03 (1.395)	12.09 (1.155)	7.172 (0.666)
0.05	29.85 (3.000)	21.33 (2.153)	14.52 (1.586)

$U = 0.5$			
0.01	17.28 (1.540)	15.29 (1.562)	9.366 (0.822)
0.025	22.56 (2.514)	14.00 (1.388)	9.398 (0.874)
0.05	25.23 (2.571)	15.62 (1.752)	10.52 (1.046)

u/u^* for N = 10,000, $U = 0.1$			
s_a	Deleterious mutations strength s_d		
	0.01	0.025	0.05
0.01	27.27 (2.824)	12.75 (1.201)	8.850 (0.594)
0.025	42.33 (4.377)	22.71 (2.166)	14.97 (1.391)
0.05	61.60 (6.219)	41.26 (4.227)	31.60 (3.915)

s_a	Deleterious mutations strength s_d		
	0.000	0.001	0.005
0.01	35.29 (2.625)	36.06 (2.392)	34.95 (2.935)
0.025	58.98 (5.651)	60.37 (5.802)	52.04 (4.535)
0.05	77.56 (7.671)	81.38 (7.505)	74.36 (6.874)

$U = 0.5$			
0.01	34.79 (3.341)	26.45 (2.919)	15.20 (2.186)
0.025	36.96 (3.835)	26.41 (3.492)	14.82 (1.593)
0.05	55.01 (6.306)	27.65 (3.137)	15.32 (1.408)

u/u^* for N = 25,000, $U = 0.1$			
s_a	Deleterious mutations strength s_d		
	0.01	0.025	0.05
0.01	62.50 (5.765)	22.74 (1.799)	19.25 (1.898)
0.025	118.38 (10.76)	63.97 (6.686)	48.40 (4.428)
0.05	204.15 (20.05)	143.50 (17.59)	84.48 (9.614)

s_a	Deleterious mutations strength s_d		
	0.000	0.001	0.005
0.01	88.88 (13.20)	91.94 (9.373)	78.62 (7.295)
0.025	161.77 (16.48)	177.02 (16.20)	148.88 (14.45)
0.05	210.74 (24.98)	219.84 (18.43)	206.44 (19.85)

3.B Supplementary figures

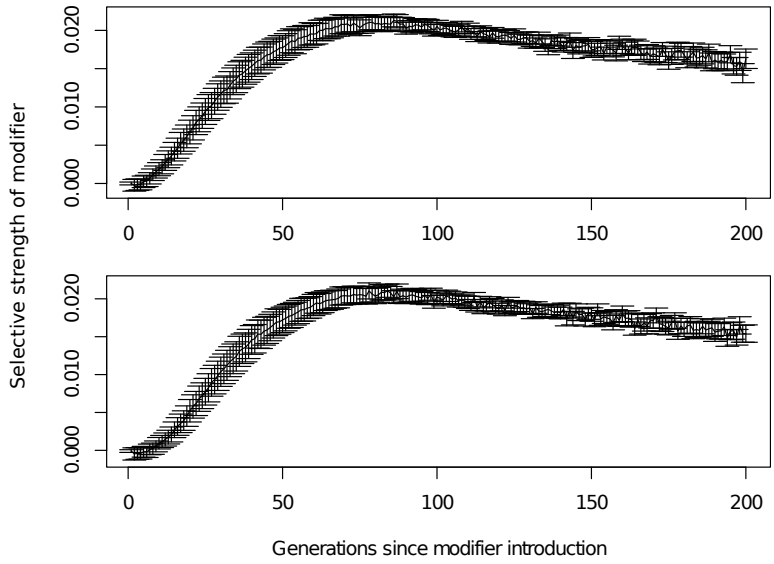


Figure 3.6: Selection acting on the modifier in the first two hundred generations, after it is introduced into a population. Both advantageous and deleterious mutants are present, with overall mutation rate U (top) or $(1 + x)U$ (bottom). Parameters are $N = 1000$, $U = 0.5$, $s_a = 0.05$, $s_d = 0.01$.

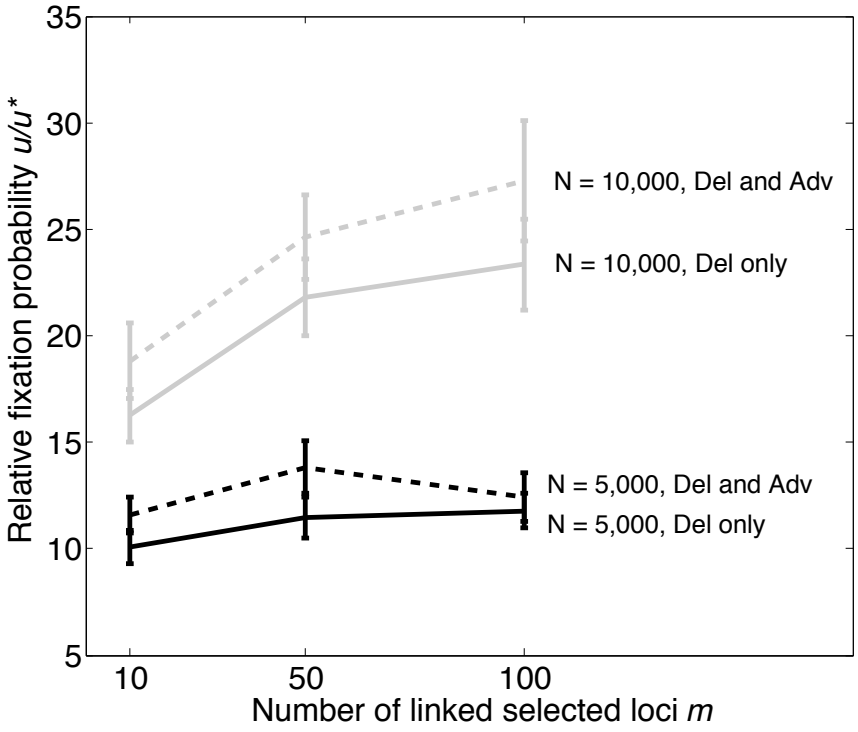
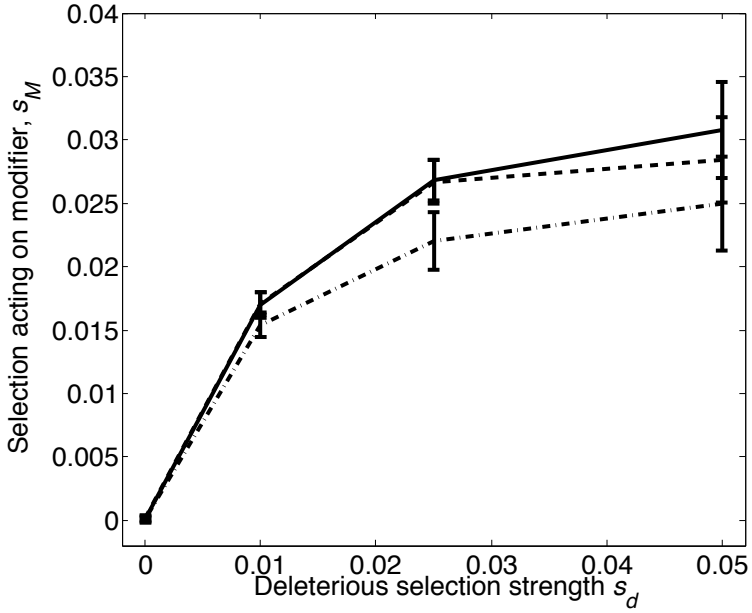
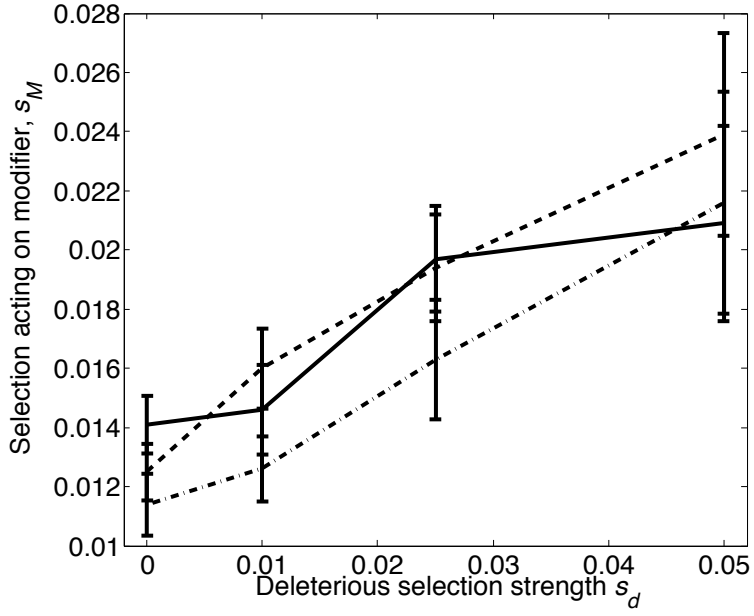


Figure 3.7: Relative fixation probability of a recombination modifier u/u^* as a function of the number of linked loci under selection. Mutations are just deleterious, or a mixture of deleterious and advantageous with strength $s_a = 0.05$. $U = 0.1$ in all cases simulated.

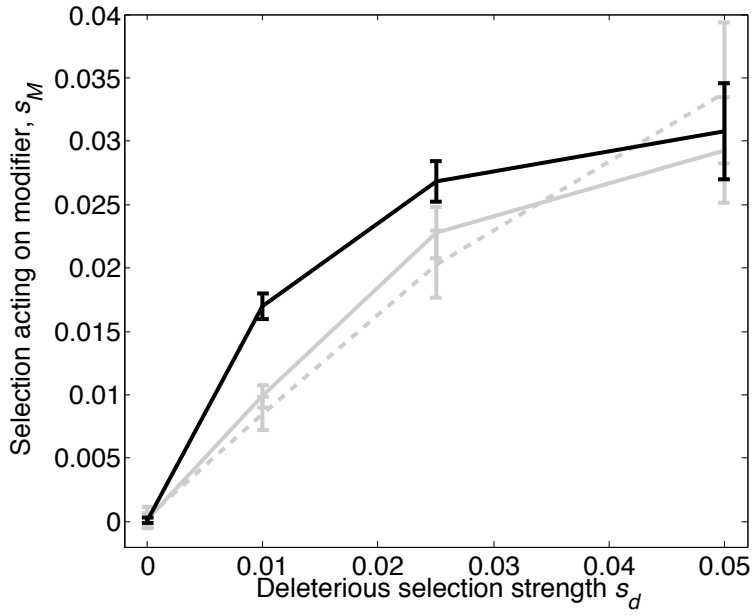


(a)

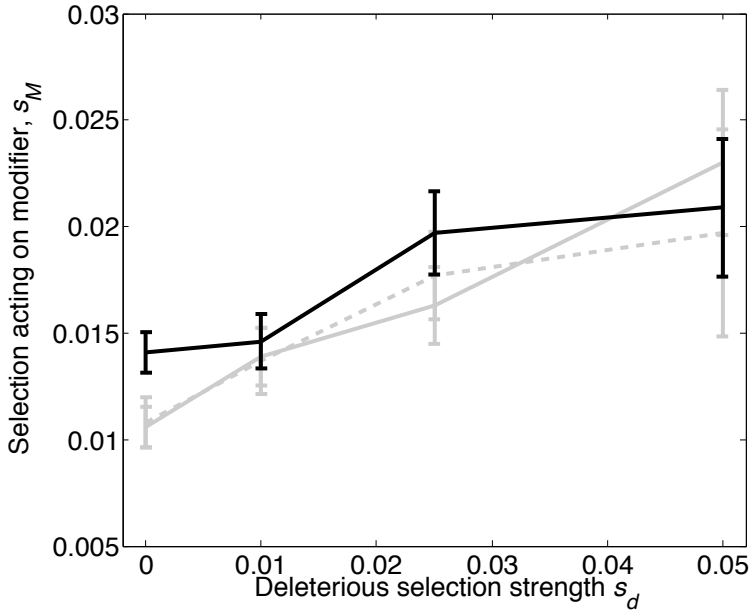


(b)

Figure 3.8: Selective advantage of a modifier, s_M, inferred for cases where mutations are deleterious only (3.8(a)) and where mutations are deleterious and advantageous (3.8(b)). The number of linked loci under selection is either 10 (dot-dashed line), 50 (dashed line) or 100 (solid line). $N = 1000$, $U = 0.5$, $s_a = 0.05$ if present.



(a)



(b)

Figure 3.9: Selective advantage of a modifier, s_M, inferred for cases where mutations are deleterious only (3.9(a)) and where mutations are deleterious and advantageous (3.9(b)). Modifier is introduced at a frequency of 5% (gray dashed line), 10% (gray solid line) or 50% (black line). $N = 1000$, $U = 0.5$, $s_a = 0.05$ if present.

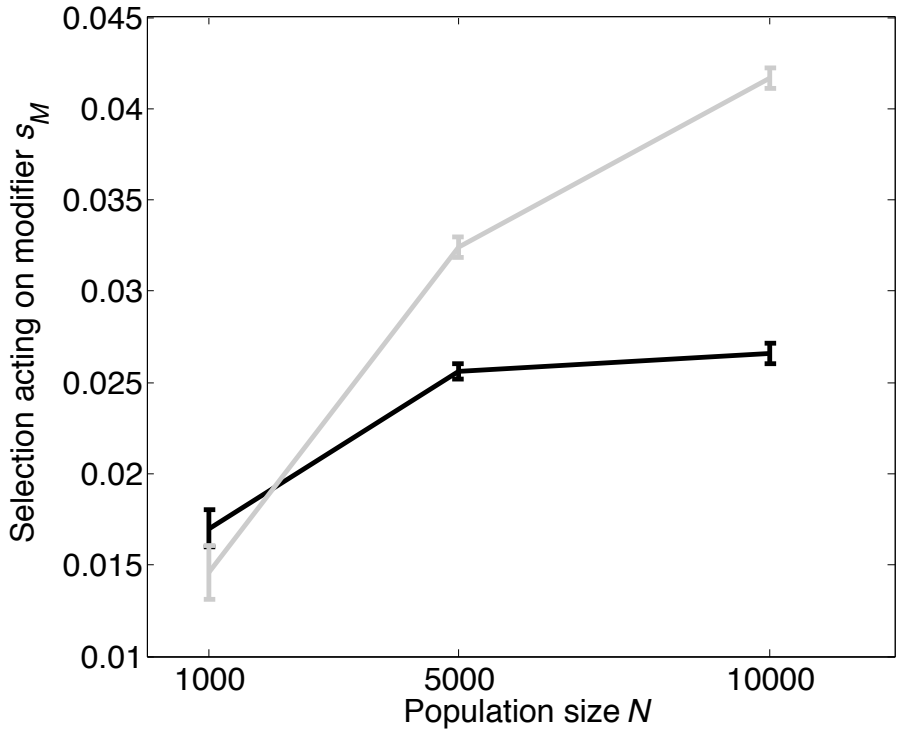
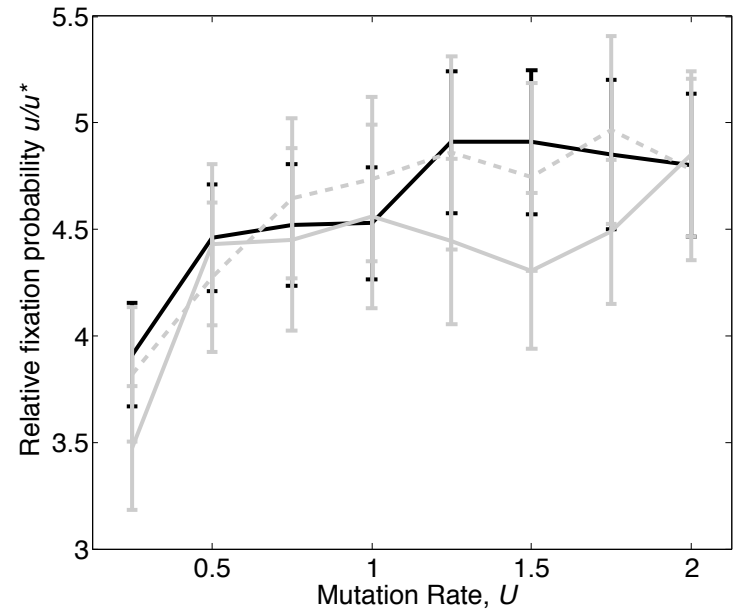
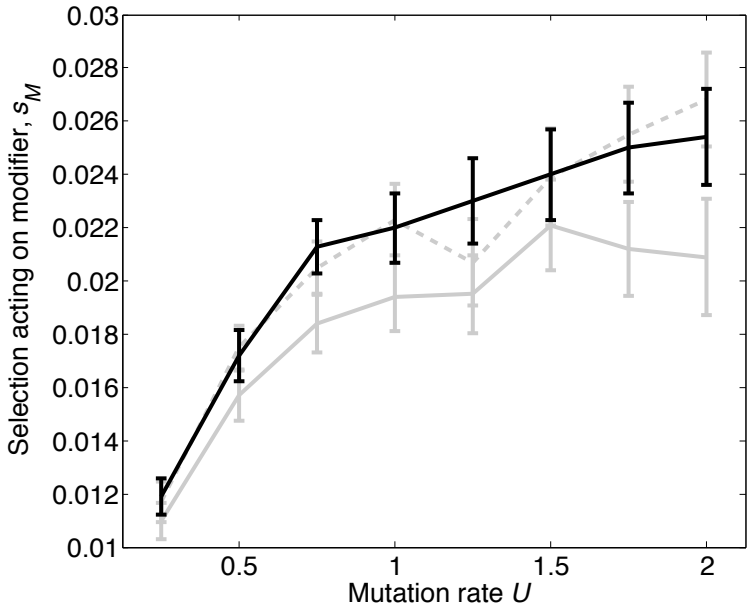


Figure 3.10: Selective advantage of a modifier, s_M as a function of population size N. Mutations are deleterious only (black line) and where mutations are deleterious and advantageous (gray line). In all cases $U = 0.5$, $s_a = 0.05$ if present.



(a)



(b)

Figure 3.11: (3.11(a)) u/u^* and (3.11(b)) s_M as a function of U if mutation is deleterious only, where there are 10 loci (gray solid line), 50 loci (grey dashed line), or 100 loci (black line) under selection. $N = 1000$, $s_d = 0.01$.

Chapter 4

Recombination and hitchhiking of deleterious alleles

Published as: Hartfield, M. and Otto, S.P., 2011. Recombination and Hitchhiking of Deleterious Alleles. *Evolution*, **65**(9): 2421–2434. Supplementary Material available on attached CD.

Chapter abstract

When new advantageous alleles arise and spread within a population, deleterious alleles at neighbouring loci can hitchhike alongside them and spread to fixation in areas of low recombination, introducing a fixed mutation load. I use branching processes and diffusion equations to calculate the probability that a deleterious allele hitchhikes and fixes alongside an advantageous mutation. As expected, the probability of fixation of a deleterious hitchhiker rises with the selective advantage of the sweeping allele and declines with the selective disadvantage of the deleterious hitchhiker. I then use computer simulations of a genome with an infinite number of loci to investigate the increase in load after an advantageous mutant is introduced. I show that the appearance of advantageous alleles on genetic backgrounds loaded with deleterious alleles has two potential effects: it can fix deleterious alleles and also facilitate the persistence of recombinant lineages that happen to occur. The latter is expected to reduce the signals of selection in the surrounding region. I consider these results in light of human genetic data to infer how likely it is that such deleterious hitchhikers have occurred in the recent evolutionary past of humans.

4.1 Introduction

The first generation of evolutionary models of advantageous alleles focused on the dynamics of single selected loci in isolation from surrounding sites (FISHER 1930; HALDANE 1927). HILL and ROBERTSON (1966) demonstrated, however, that selection acting at one site in finite populations interferes with the efficacy of selection at surrounding sites, hampering the spread of neighboring beneficial alleles, even in the absence of fitness interactions among the sites. As pointed out verbally by FISHER (1930), and subsequently by HILL and ROBERTSON (1966), CHARLESWORTH *et al.* (1993) and generalized by RICE (1999), selection on linked sites reduces the effective number of lineages contributing to future generations to those lineages with the highest fitness. Such genetic bottlenecks (also called 'Hill-Robertson' interference) increase the contribution of drift relative to selection, such that advantageous alleles are less likely to spread and they spread more slowly than predicted by their direct effects on fitness. In a general analysis by BARTON (1995b), this interference was shown to reduce the fixation probability of beneficial alleles linked to other selected sites. Breaking down interference among selected loci has also been shown to favor increased rates of sex and recombination (OTTO and BARTON 1997; BARTON and OTTO 2005; ROZE and BARTON 2006).

In addition to affecting neighbouring loci under selection, MAYNARD SMITH and HAIGH (1974) showed that the dynamics of a single selected locus impacts surrounding neutral loci. In particular, an advantageous allele sweeping through a population reduces, on average, the genetic variance around the site of it (see also THOMSON (1977)). This phenomenon provides a mechanism for detecting regions experiencing selection, forming the basis for the HKA test (HUDSON *et al.* 1987), for example.

Relatively little attention has been paid, however, to the effect that selection on neighboring sites might have on the net fitness change associated with the fixation of a focal beneficial allele and on the patterns of variation at surrounding selected sites

(see, for example, YU and ETHERIDGE (2010) regarding beneficial alleles segregating in the background, and HADANY and FELDMAN (2005) regarding deleterious alleles in the background). In this chapter, I consider a focal site carrying a new beneficial allele in the presence of neighboring sites subject to deleterious mutations. I calculate the chance that a linked deleterious allele hitchhikes to fixation along with the beneficial allele, as a function of the rate of recombination between them, and describe the implications for patterns of variation expected within the region of a selective sweep. This work builds upon a recent simulation study by HADANY and FELDMAN (2005), as well as complementary analytical work for asexual organisms (JOHNSON and BARTON 2002; BACHTROG and GORDO 2004; YU and ETHERIDGE 2008; YU *et al.* 2010). Specifically, HADANY and FELDMAN (2005) demonstrated that beneficial alleles sweeping to fixation in a purely asexual population often carry along linked deleterious alleles. The fixation of deleterious alleles by hitchhiking generates a fixed mutation load that must await a future adaptive sweep by a back or compensatory mutation in order for it to be erased. This work provides an analytical prediction of the probability of such undesirable hitchhikers, allowing for arbitrary rates of recombination between the sites under selection.

4.1.1 Empirical Background

Recent studies of amino-acid substitution data suggest that advantageous mutants are present at higher rates then previously assumed. Although precise values remain a matter of debate (EYRE-WALKER 2006), BIERNE and EYRE-WALKER (2004) estimated that approximately 45% of amino acid substitutions are adaptive in *Drosophila melanogaster*, equating to one substitution, on average, every 450 generations. Later studies have found that between 30% – 60% of substitutions in *Drosophila melanogaster* coding and non-coding regions are adaptive (ANDOLFATTO 2005; OBBARD *et al.* 2009; ANDOLFATTO 2007; SHAPIRO *et al.* 2007), highlighting the prevalence of beneficial

mutations. Similar values have been observed in the wild mouse *Mus musculus castaneus* (HALLIGAN *et al.* 2010). In hominids, studies suggests that this rate is lower; BOYKO *et al.* (2008) and EYRE-WALKER and KEIGHTLEY (2009) found that on average 5% of amino-acid substitutions were adaptive if recent population bottlenecks were taken into account.

Another method to detect the presence of advantageous mutations is through investigating the underlying distribution of fitness effects among mutations. Using such a method SHAW *et al.* (2002) suggested that half of all mutations in *Arabidopsis thaliana* increased fitness (although see KEIGHTLEY and LYNCH (2003)). Even in fairly adapted lab strains of *Saccharomyces cerevisiae*, JOSEPH and HALL (2004) estimated that around 5.75% of spontaneous mutations were beneficial (HALL and JOSEPH 2010).

The strength of selection acting on beneficial alleles is also subject to much debate and is expected to depend on the nature of past environmental changes, both biotic and abiotic (ELENA and LENSKI 2003). On the lower end, JENSEN *et al.* (2008) estimated that advantageous mutants have had a mean selection coefficient of $s_a \approx 10^{-4}$ in *Drosophila*. On the upper end, very large selection coefficients have been detected in experimental evolution studies with bacteria and viruses, with an average $s_a \approx 2$ found in *Pseudomonas fluorescens* exposed to a novel carbon source (BARRETT *et al.* 2006) and s_a ranging between 6 and 14 in the bacteriophage ϕX174 subjected to heat stress (BULL *et al.* 2000). However, these strongly-selected beneficial mutations may not be representative of the mean selective strength.

While there is increasing evidence for the frequent spread of advantageous alleles, it is an inescapable fact that most spontaneous mutations that affect fitness are deleterious (CROW 1970) and are maintained in populations at a low frequency by recurrent mutation (WRIGHT 1931). These mutation rates can be substantial; for example, the per-generation genomic deleterious mutation rate U_d in *Drosophila* has been estimated at 1.2 (HAAG-LIAUTARD *et al.* 2007; KEIGHTLEY *et al.* 2009), with estimated rates of

U_d of around 4.2 in hominids (EÖRY *et al.* 2010). Deleterious mutation rates are lower in microbes, however. In non-mutator strains of yeast, HILL and OTTO (2007) estimated $U_d = 0.013$ for mutations acting on sporulation ability and $U_d = 0.0003$ for those affecting growth rate.

If selection acts against deleterious mutations with a selection coefficient of s_d, then it is expected that a total of $\sim U_d/s_d$ mutations segregate within a population at mutation-selection balance (ignoring genetic associations among them). Even when U_d is less than one, the expected number of deleterious mutations carried by an individual may be much greater than one. Consequently, newly arisen advantageous alleles may occur within chromosomes also bearing deleterious alleles nearby. In the next section, I develop a model that describes the fate of a deleterious mutation that occurs in the genetic background of a novel beneficial allele. I later return to estimates of the rates of mutation and selection coefficients to assess how likely it is that deleterious alleles hitchhike to fixation, and how this depends on the mode of reproduction and the effective rate of recombination within a species.

4.2 Semi-Deterministic Model

I first present a semi-deterministic calculation of the fixation probability of a haplotype carrying both an advantageous and a deleterious allele using classic population genetics. In the next section, I build a stochastic diffusion model of the appearance and spread of this haplotype, but the calculations presented in this section help to develop an understanding of the key forces at work and so are a natural first step in investigating this problem.

I consider a finite population of N haploid chromosomes with discrete generations, using a standard Wright-Fisher model (FISHER 1930; WRIGHT 1931). I am interested in the dynamics of a newly arisen beneficial allele at a locus A. The genome in which

A first arises may carry one or more deleterious alleles. Deleterious alleles that are only loosely linked to locus A are unlikely to rise substantially in frequency and are ignored. I focus only on the single most closely linked of these deleterious mutations and call this second locus B, with recombination between A and B occurring at rate r. At locus A, the advantageous allele A_1 has a selective advantage s_a over the wild-type allele A_0. At locus B, the deleterious allele B_1 is selected against with selection coefficient s_d, relative to the wild-type allele B_0. As I am investigating the fixation probability of a weakly-selected deleterious allele as it hitchhikes with an adaptive allele, I assume $s_a > s_d$, so that the advantageous-deleterious haplotype has a net beneficial effect, $s_{net} = s_a - s_d$. For clarity of presentation, I assume additive selection, so all of the analytical results continue to apply if s_d is replaced by $s_{net} - s_a$, wherever it occurs.

For each haplotype, I write a 0 subscript if the wild-type allele is present at the locus and a 1 subscript otherwise, in the order AB. All possible haplotypes, along with their fitnesses, are given in table 4.1. In particular, the advantageous-deleterious haplotype is denoted A_1B_1, and when this haplotype first appears, the remainder of the population is either A_0B_0 (wildtype) or A_0B_1 (bearing the deleterious allele). The latter haplotype (A_0B_1) is assumed to be rare and is ignored in the following analysis to simplify the calculations; simulations described in a later section indicate that this assumption introduces little bias. I also assume that no further mutation occurs at either of the loci during the course of the sweep, although the model can be modified to take this into account.

Table 4.1: Table of haplotypes.

Haplotype	Fitness, w	Locus 1	Locus 2
$A_0\ B_0$	1	Wildtype	Wildtype
$A_1\ B_0$	$1 + s_a$	Beneficial	Wildtype
$A_0\ B_1$	$1 - s_d$	Wildtype	Deleterious
$A_1\ B_1$	$1 + s_a - s_d$	Beneficial	Deleterious

Let $p(t)$ denote the frequency of the A_1B_1 haplotype, where t is the number of generations since the beneficial allele arose and p_0 is its initial frequency (generally $1/N$). When the A_1B_1 haplotype first arises, it becomes established within the population with a probability u that is approximately twice the net selection coefficient, $2s_{net}$ (HALDANE 1927). It is further assumed that $s_{net} \ll 1$ and that the population size is large (see next section for results that apply in smaller populations).

In the following derivation, I only consider those A_1 alleles that survive stochastic loss while rare. Once established, the frequency of A_1B_1 can be modelled by the standard deterministic equation for haploid selection (HALDANE 1924):

$$p(t) = \frac{p_0(1+s_{net})^t}{p_0(1+s_{net})^t + 1 - p_0} \tag{4.1}$$

which assumes that s_{net} is constant over time; that is, recombination has a negligible effect on the frequency of the A_1B_1 haplotype, if it does not result in the production of an A_1B_0 haplotype that goes to fixation. Among those alleles that succeed in fixing, the trajectory of the A_1B_1 haplotype is slightly faster, on average, than given by equation 4.1 as the advantageous haplotype is sampled more frequently than expected (MAYNARD SMITH and HAIGH 1974; BARTON 1994). This initial acceleration is taken into account in the diffusion model developed below; it turns out to have little effect, however, because the recombination events that break apart the A_1B_1 haplotype are most likely to occur when the A_1B_1 haplotype is intermediate in frequency and not when it initially occurs.

The goal is to calculate the probability, P, that the A_1B_1 haplotype is not broken apart by recombination before the advantageous A_1 allele fixes within the population. If such a recombination event has not yet occurred, there are approximately $p(t)$ of the A_1B_1 haplotypes and $1-p(t)$ of the A_0B_0 haplotypes (ignoring the rare A_0B_1 individuals), so that matings between these two haplotypes occur at frequency $2p(t)(1-p(t))$. Among the offspring of these matings, r will be recombinant, half of which will carry

the most fit A_1B_0 haplotype and half of which will carry the least fit A_0B_1 haplotype. Even when once produced, the most fit recombinant may fail to establish itself within the population due to chance loss while rare. In Appendix 4.A, I use branching processes to show that the probability that a single new A_1B_0 haplotype establishes within the population if it appears at time t equals:

$$\Pi(t) = \frac{2s_a s_d}{s_a p(t) + s_d(1 - p(t))} + O(s^2). \tag{4.2}$$

The derivation of equation 4.2 takes into account the fact that the A_1B_0 haplotype has fitness $1+s_a$ relative to the population mean fitness $1+p(t)(s_a-s_d)$, which is changing over time according to equation 4.1. As expected, if the A_1B_0 recombinant haplotype arises while $p(t) \approx 0$, the recombinant lineage will establish with probability nearly equal to $2s_a$, the fixation probability of an advantageous A_1 allele in an otherwise wild-type population. Also as expected, if the A_1B_0 recombinant haplotype arises when $p(t) \approx 1$, the recombinant lineage will establish with probability nearly equal to $2s_d$, the fixation probability of a haplotype that has shed the deleterious allele B_1 in a population that otherwise carries both A_1 and B_1. I call A_1B_0 haplotypes that succeed in establishing while rare "successful recombinants".

Altogether, $\kappa(t) = rp(t)(1-p(t))\Pi(t)$ is the probability that an A_1B_0 recombinant haplotype appears at time t and goes on to establish within the population. Note however that this calculation does not specify whether the A_1 or B_0 allele will fix first; in many cases, if a recombinant appears and fixes with probability $\Pi(t)$, the actual fixation of the A_1B_0 haplotype would occur after A_1 has reached fixation.

To calculate the overall probability, P, that the A_1B_1 haplotype is never broken apart by recombination, I must calculate the probability that in every generation, t, none of the N offspring are successful recombinants. Assuming weak selection, such that both $\Pi(t)$ and $\kappa(t)$ are small, the probability that a deleterious hitchhiker will be carried to fixation by the spread of a linked beneficial allele is given by:

$$
\begin{aligned}
P &= \prod_{t=0}^{\infty}(1-\kappa(t))^N \\
&\approx \prod_{t=0}^{\infty} exp[-N\kappa(t)] \\
&= exp\left[\sum_{t=0}^{\infty} -N\kappa(t)\right] \\
&\approx exp\left[\int_{t=0}^{\infty} -N\kappa(t)dt\right] \qquad (4.3)
\end{aligned}
$$

Overall, P gives the probability that a fitter recombinant never establishes, assuming that the A_1B_1 haplotype is not lost stochastically when it first appears. The probability that the A_1B_1 haplotype succeeds in establishing initially and fixing within the population is thus u $(= 2s_{net})$ times P. This equation is analogous to equation 16 in YU and ETHERIDGE (2010), who used a Moran model to estimate the fixation probability of two competing beneficial mutations, with recombination between the two loci.

Equation 4.3 can be solved by integrating over the allele frequency dynamics rather than over time and replacing the integral with

$$\int_{p=p_0}^{1} -\frac{N\ \kappa(p)}{dp/dt}dp \qquad (4.4)$$

In this haploid model with weak selection, and negligible effect of recombination, $dp/dt = (s_a - s_d)p(1-p)$. Carrying out the integration, the probability that a fitter recombinant never establishes is given by:

$$P \approx exp\left[-\frac{2\ N\ r\ s_a\ s_d\ \ln(s_a/s_d)}{(s_a - s_d)^2}\right]$$

where p_0 was assumed negligible relative to terms on the order of one. At this point, I

can eliminate the population size from the result by measuring the scaled selection and recombination rates within the population, defined as $S_d = Ns_d$, $S_a = Ns_a$, $S_{net} = N(s_a - s_d)$, and $\rho = Nr$:

$$P \approx \left(\frac{S_a}{S_d}\right)^{-\omega} \tag{4.5}$$

where ω is the compound parameter defined by

$$\omega = 2\ \rho\ \frac{S_a\ S_d}{S_{net}^2} \tag{4.6}$$

The hitchhiking process thus depends primarily on these scaled parameters and not separately on the population size and selection or recombination parameters. Equation 4.5 shows that the probability of hitchhiking to fixation declines exponentially with the recombination rate between the loci and with the number of individuals within the population. The probability of hitchhiking is especially small when the strength of selection for the beneficial allele and against the deleterious allele are similar (S_{net} small), as this will cause the sweep of the A_1B_1 haplotype to take longer and allow for more recombination events.

To determine how small the recombination rate must be in order for hitchhiking to occur with a particular probability of interest, c, I set $P = c$ and solve for ρ:

$$\rho_{crit} = \frac{S_{net}}{S_d}\left[\frac{\ln(\frac{1}{c})}{2(1+\frac{S_d}{S_{net}})\ln(1+\frac{S_{net}}{S_d})}\right]. \tag{4.7}$$

This gives us the recombination rate below which hitchhiking to fixation will occur with frequency greater than c, as a function only of the scaled selection coefficients S_d and S_{net}. At this point, I hold off discussing these results further until the next section, where I derive a stochastic solution.

4.3 Stochastic Model

The above analysis assumes that the population is very large, allowing us to combine stochastic results for the establishment of particular haplotypes while rare, with deterministic equations for the spread of these haplotypes. The above does not, however, take into account chance fluctuations in haplotype frequencies or the initial acceleration caused by considering only those trajectories where the beneficial allele becomes established (MAYNARD SMITH and HAIGH 1974; BARTON 1994; OTTO and BARTON 1997; DESAI and FISHER 2007). To account for these effects, I now derive a stochastic solution for this problem.

Again ignoring the rare deleterious-only lineage, I model the change in frequency, $p(t)$, of the A_1B_1 haplotype using a diffusion approximation. If a successful recombinant appears, however, the diffusion process is killed. As described by KARLIN and TAYLOR (1981), the probability that the process is not ultimately killed, $P(p)$, given that A_1B_1 is currently at frequency p, satisfies:

$$\frac{1}{2}V(p)\frac{d^2P(p)}{dp^2} + M(p)\frac{dP(p)}{dp} - K(p)P(p) = 0 \tag{4.8}$$

where $M(p)$ is the mean change in p over a timestep measured in N generations; $V(p)$ is the variance in change of p; and $K(p)$ is the killing function, which denotes the probability of the process being 'killed' while the A_1B_1 haplotype is at frequency p. In this model, killing occurs if recombination forms a fitter haplotype (i.e., A_1B_0) that succeeds in establishing itself within the population. To solve equation 4.8, I use the boundary conditions $P(0) = P(1) = 1$; that is, the system cannot be killed if either the A_1B_1 or A_0B_0 haplotype is fixed. Further descriptions of similar diffusion models with killing are available in KARLIN *et al.* (1967) and section 15.10 of KARLIN and TAYLOR (1981); in particular, a related model is described where the diffusion process is killed whenever any recombinant is formed (A_1B_0 or A_0B_1), regardless of whether

the recombinant succeeds in establishing itself.

As with standard diffusion models investigating an allele under weak directional selection in a haploid population (KIMURA 1970; EWENS 2004), I obtain the values $M(p) = S_{net}\ p(1-p)$ and $V(p) = p(1-p)$, where $S_{net} = N(s_a - s_d)$ (see Supplementary Material, section 2, available on the attached CD). The killing term is obtained by taking the probability that the process is killed in a particular generation, $1-(1-\kappa)^N \approx N\kappa = N\ r\ p(1-p)\ \Pi$, and scaling in such a way that the killing term remains finite over the timestep of N generations, as $N \to \infty$ (KARLIN and TAYLOR 1981). By doing so, I obtain the killing function $K(p) = \rho\ p(1-p)\ \pi(p)$, where $\rho = Nr$ and $\pi(p)$ is the scaled version of the establishment probability of the A_1B_0 recombinant, Π (Equation 4.2):

$$\pi(p) = \frac{2\ S_d\ (S_{net} + S_d)}{p\ S_{net} + S_d} \tag{4.9}$$

The diffusion approximation assumes that S_{net}, S_d, and ρ remain finite as $N \to \infty$.

Plugging these diffusion coefficients into equation 4.8 and dividing by $p(1-p)$, the probability that the process is not killed, $P(p)$, given the current frequency p satisfies:

$$\frac{1}{2}\frac{d^2P(p)}{dp^2} + S_{net}\frac{dP(p)}{dp} - \rho\ \pi(p)\ P(p) = 0 \tag{4.10}$$

If the process is not killed, there are two potential outcomes: fixation of A_0B_0 or fixation of A_1B_1. If I wish to know the probability that a particular advantageous allele that succeeds in fixing carries along with it a deleterious allele, I must rederive the diffusion model conditional on A_1 establishing within the population. In Appendix 4.B, I show that the conditional probability $P^*(p)$ that the process is not killed (i.e., the deleterious allele B_1 fixes) among those cases where A_1 sweeps to fixation satisfies:

$$\frac{1}{2}\frac{d^2P^*(p)}{dp^2} + S_{net}\ \frac{1+e^{-2pS_{net}}}{1-e^{-2pS_{net}}}\frac{dP^*(p)}{dp} - \rho\ \pi(p)\ P^*(p) = 0. \tag{4.11}$$

The differential equations 4.10 and 4.11 were solved in *Mathematica* 6.0 (Supplementary Materials), yielding the somewhat cumbersome equations 4.23 and equation (4.24), respectively. These can be solved numerically for the probability that the process is not ultimately killed (i.e., the probability that a successful recombinant never appears).

$P^*(p_0)$ as given by 4.24 is the main quantity of interest in this chapter. It describes the probability that an A_1 allele that fixes within a population carries along with it a linked deleterious allele B_1, given that the initial frequency of the A_1B_1 haplotype is p_0. Although equations 4.23 and 4.24 should be used in any numerical analysis, further insight is provided by approximating $P^*(p_0)$ as an exponentially decreasing function of the recombination rate (as in the semi-deterministic analysis). Assuming that selection is strong relative to drift ($S_d, S_{net} \gg 1$), that the frequency of the A_1B_1 haplotype when the A_1 allele first appears is negligibly small ($p_0 \ll 1$), and that linkage is not too frequent ($\rho \ll S_d, S_{net}$), I obtain:

$$P^*(p) \approx \left(e^{-1/S_d} \, \frac{S_a}{S_d} \right)^{-\omega} \tag{4.12}$$

(see details in Supplementary Materials, section 3). Again, this can be used to calculate a critical value of recombination above which hitchhiking is unlikely to occur. Specifically, I solve equation 4.12 for the rate of recombination necessary for the deleterious B_1 allele to fix with probability c, given that the beneficial allele A_1 initially appears with B_1 and ultimately fixes:

$$\rho_{crit} = \frac{S_{net}}{S_d} \left[\frac{\ln(\frac{1}{c})}{2(1 + \frac{S_d}{S_{net}})(\ln(1 + \frac{S_{net}}{S_d}) - 1/S_d)} \right]. \tag{4.13}$$

For example, when $c = 1/2$, the term in square brackets is approximately 1/4 as long as neither S_d nor S_{net} is too small (see the figure in section 3 of the Supplementary Materials). Thus, as a rough rule of thumb (using unscaled parameters), the recombination

rate r must be less than $1/4$ of $s_{net}/(Ns_d)$ for there to be at least a 50% chance that the deleterious allele hitchhikes to fixation.

Hitchhiking events are thus likely to occur over larger regions of the genome if the net selection coefficient acting on the A_1B_1 haplotype, s_{net}, is stronger, because sweeps then occur faster. Conversely, the stronger the disadvantage of the deleterious allele, s_d, the less likely a hitchhiker will fix, because recombinant A_1B_0 haplotypes are so much more fit. Finally, the larger the population size, the less likely that a hitchhiker will fix, simply because there are more individual chances for recombination to occur while the population remains polymorphic.

These patterns are illustrated in Figure 4.1, which gives the probability that the deleterious B_1 allele hitchhikes to fixation given that the beneficial A_1 allele fixes, with darker shading corresponding to higher probabilities. These contour plots are based on the exact solution 4.24 to the diffusion equation for $P^*(p)$. The thick dashed curves show the approximate equation 4.13 for the critical value of the recombination rate, ρ, below which I expect deleterious alleles to hitchhike to fixation more than c of the time ($c = 10\%$, 50%, or 90%) when they occur on the haplotype bearing a new beneficial allele; these curves accurately follow the appropriate contour lines as long as selection is not too weak (roughly, $S_{net}, S_d \geq 2$).

4.3.1 Comparison to the case of a linked neutral allele

The dynamics of neutral loci are likely to be affected by the spread nearby of a beneficial allele whenever r is approximately less than s_a (MAYNARD SMITH and HAIGH 1974). This rule cannot be used to compare to equation 4.13 directly, however, because the criteria for being "affected" is now quite strict: the linked B_1 allele must fix due to the sweep. I thus briefly describe a corresponding model for the case when B is neutral (full details are provided in the Supplementary Materials, section 4).

The diffusion equations remain essentially the same, except that the killing term

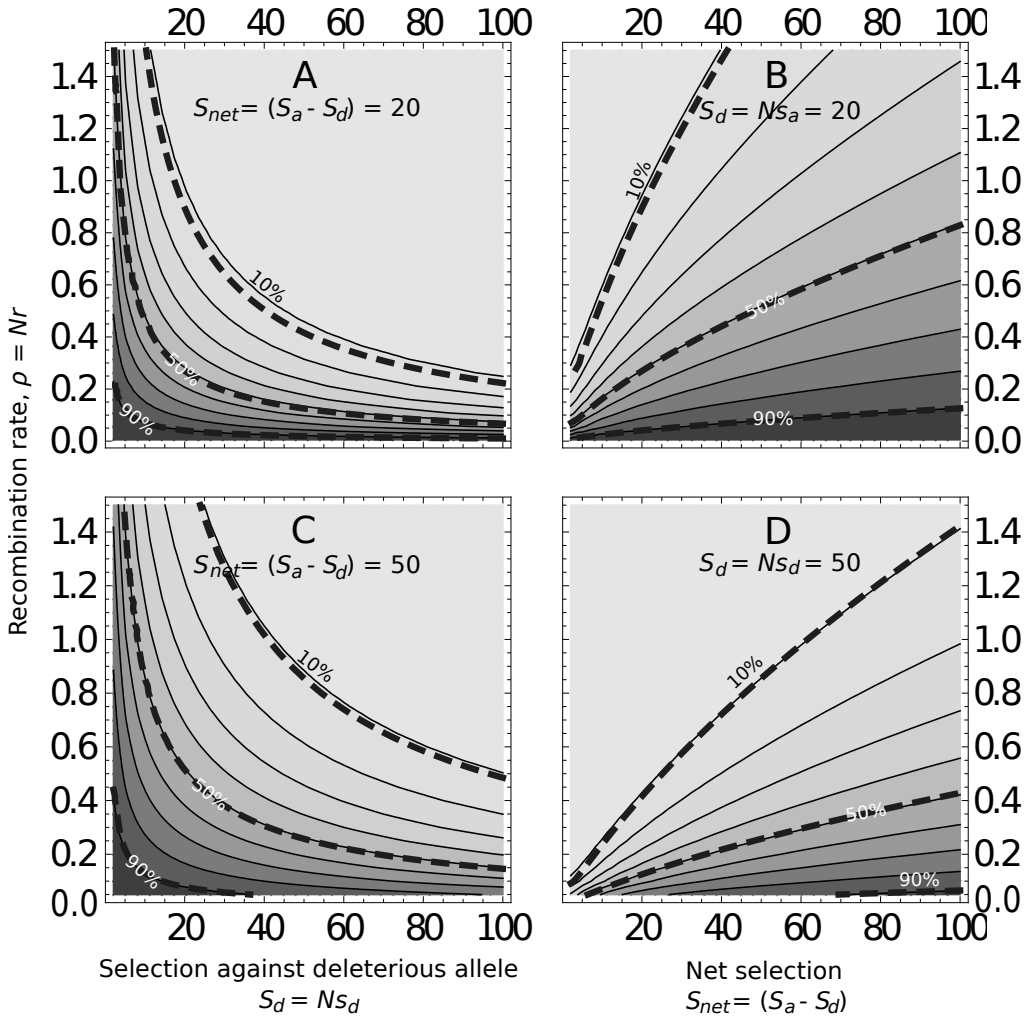


Figure 4.1: Contour plots of the fixation probability of the deleterious B_1 allele, given that the A_1B_1 haplotype appears initially at frequency of $1/N$ and that the A_1 allele is not lost stochastically (10% contour intervals based on equation 4.24). The graphs are shown for N = 10,000, although the results are not very sensitive to N, as long as the scaled parameters are held constant. In each case, ρ is plotted along the y-axis versus S_d along the x-axis (left panels) with S_{net} = 20 (top) or 50 (bottom) or versus S_{net} along the x-axis (right panels) with S_d = 20 (top) or 50 (bottom). The dashed curves show the predicted thresholds below which there is a greater than $c = 10\%$, 50%, and 90% probability of hitchhiking, based on equation 4.13; in each case this threshold coincides closely with the appropriate contours.

must be revised now that the recombinant A_1B_0 haplotype is no more fit than the A_1B_1 haplotype that is spreading through the population. I assume that, whenever a recombinant A_1B_0 haplotype appears, the probability that this haplotype becomes the ancestor of the population at some distant future point in time is very nearly $1/(Np)$. This assumes that any individual carrying the A_1 allele alive at that time is equally likely to be the lucky one to ultimately fix and give rise to the entire descendant population. Using $1/(Np)$ instead of Π for the fixation probability of the recombinant A_1B_0 haplotype, I obtain the revised killing function, $K(p) = \rho\ p(1-p)\ 1/p$, for use in the diffusion equation 4.8, assuming that allele A_1 fixes. The conditional probability of the process not being killed was then obtained using *Mathematica* 6.0.

Focusing on the conditional probability that the process reaches fixation on A_1 before being killed by the appearance of a successful recombinant, I again obtained an approximation assuming that selection is strong relative to drift:

$$P^*(p_0) = (2\ e^{\gamma}\ S_{net})^{-\rho/S_{net}}. \tag{4.14}$$

where $\gamma = 0.577$ is Euler's constant. I have persisted in referring to the net selection on the A_1B_1 haplotype as S_{net} despite the fact that now $S_{net} = S_a$ for ease of comparison with the previous case.

Again, solving this equation for the critical value of ρ below which hitchhiking to fixation occurs more than a proportion c of the time, I get:

$$\rho_{crit}^{neutral} = S_{net}\left[\frac{\ln(\frac{1}{c})}{\gamma + \ln(2\ S_{net})}\right]. \tag{4.15}$$

For $c = 1/2$, the term in square brackets is approximately 1/4 when $S_{net} = 5$, and it continues to decline (but slowly) as S_{net} increases. Thus, as a rough rule of thumb, r must be less than $\approx 1/4$ of s_{net} for there to be a 50% chance that a neutral allele hitchhikes to fixation. Again, such hitchhiking events are likely to occur over larger regions

of the genome when the sweeps are faster (s_{net} large). The key difference, however, from the case with a deleterious hitchhiker is the absence of Ns_d in the denominator of this rule, which makes it easier to satisfy than the case of a deleterious hitchhiker (assuming selection is strong relative to drift). Figure 4.2 shows just how much more likely it is for alleles at locus B to hitchhike to fixation along with allele A when the B locus is neutral (thick top curve) than when it is subject to selection against deleterious mutations (dashed curves).

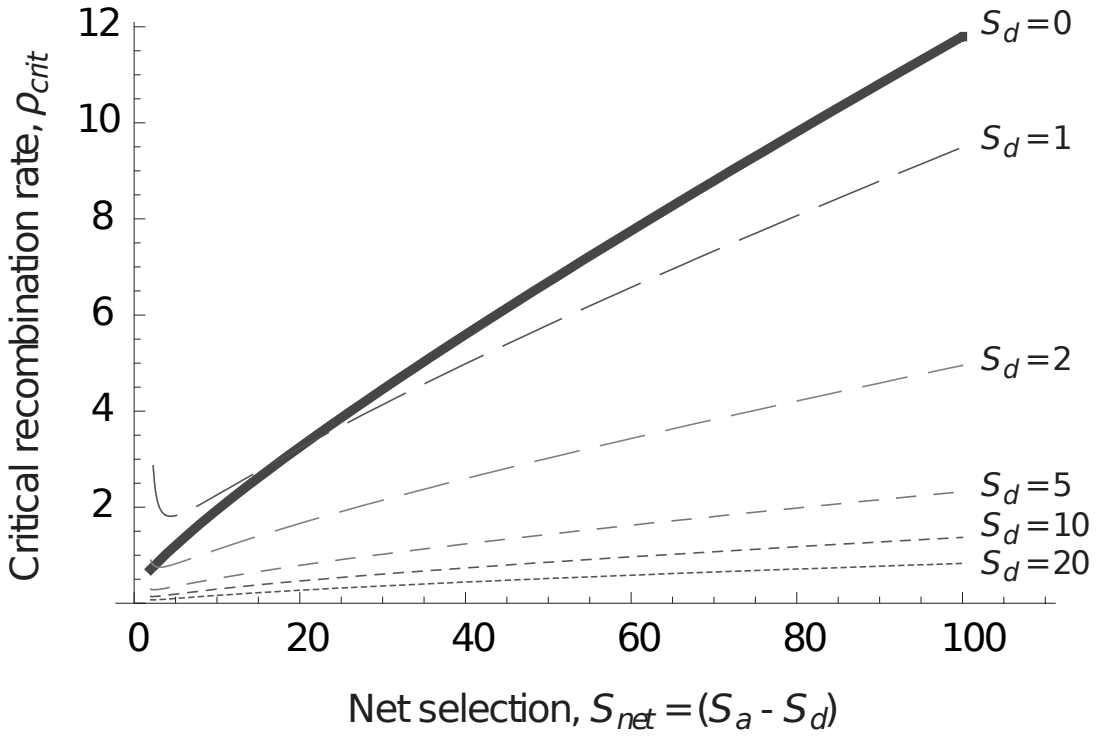


Figure 4.2: The critical value of the recombination rate, ρ_{crit}, below which there is a greater than $c = 50\%$ probability that the deleterious B_1 allele will hitchhike to fixation along with the advantageous allele, as a result of their initial association based on the approximate equations 4.13 and 4.15. In each case, ρ_{crit} is plotted along the y-axis versus S_{net} along the x-axis, for varying values of S_d. The case of a neutral linked allele at the B locus is given by the thick top curve. (The upturns in some of the curves near the origin as well as crossing of some of the curves are caused by inaccuracies in these approximations when selection is weak relative to drift.)

The fact that neutral alleles are much more likely to hitchhike to fixation than linked deleterious alleles has another important implication. Namely, the presence of a linked deleterious allele increases the chance that surrounding genetic variation will be rescued by recombination. Had there been no linked sites under selection, I would expect a region surrounding a sweep to be entirely fixed when $\rho < \rho_{crit}^{neutral}$ in the majority of cases (equation 4.15). If a beneficial allele first occurs on a chromosome containing a deleterious allele, however, this region is greatly reduced to $\rho < \rho_{crit}$ (equation 4.13), as illustrated in Figure 4.2. Consequently, linkage to sites carrying deleterious alleles reduces the impact of selective sweeps, making it less likely that surrounding genetic variation will be lost.

Turning this argument around, a recently fixed beneficial allele might have been strongly selected, but appear to have been weakly selected based on the amount of genetic variation remaining in the region. This is because recombinants were favoured that untied the beneficial allele from the deleterious genetic baggage with which it arose. Furthermore, I would expect that genetic variation should more often be rescued by the appearance of more fit recombinants on the side of a selective sweep that bears a higher density of other sites under selection. In Appendix 4.C I simulate a three-locus model with one locus subject to advantageous mutation, one locus being a neutral marker, and one locus subject to recurrent deleterious mutation, with the beneficial mutant is placed on a randomly selected genetic background. As confirmed in Figure S.1 the sweep of neutral diversity is less severe in cases where selection acts on the locus subject to deleterious mutations.

4.3.2 Two competing beneficial mutations

The above analyses can also be used to solve a related problem of beneficial mutations competing for fixation in the presence of recombination, as considered by YU and ETHERIDGE (2010). If a beneficial allele is rising in frequency when a second beneficial

allele appears at a linked site, then it is possible for the first beneficial allele to be lost if the second allele is more strongly favored and if a recombinant that brings together both alleles onto the same haplotype fails to establish in time.

Although technically there are three chromosome types to be considered before the recombinant appears (00, 10, and 01, where the “1" now indicates a beneficial mutation at the first and second sites), I can approximate this scenario as did YU and ETHERIDGE (2010) by assuming that the 00 wildtype is rapidly eliminated, so that the frequencies of 01 and 10 sum roughly to one. This approximation performs surprisingly well for this problem because most of the recombination events do not occur until the 10 and 01 haplotypes are both common.

Equation 4.1 then describes the spread of the more fit 01 haplotype, whose frequency is $\approx p(t)$ (frequency of 10 $\approx 1 - p(t)$), with s_{net} equal to the difference in fitness between 10 and 01 individuals. Equation 4.2 describes the fixation probability of a recombinant double mutant, with s_a and s_d giving the selective advantage of the double mutant when it appears in a population predominantly composed of 10 and 01 individuals, respectively. All of the subsequent results described above then follow. Figure 4.7 shows that equations 4.5 and 4.12 provide an excellent estimate of the probability that recombination successfully rescues both beneficial mutations. Although similar in spirit to the work of YU and ETHERIDGE (2010), this analyses have the advantage of providing closed form solutions that appear to accurately capture the stochastic nature of recombination rescuing combinations of beneficial alleles at two selected loci.

4.4 Two-locus simulations

In order to investigate the accuracy of the above results, I compare both the semi-determinstic and stochastic models to Monte Carlo simulations. Simulations start with a population of N haploid chromosomes, each consisting of two linked loci. Fitness is

assumed to be additive.

An initial proportion p_0 of the population is assigned the advantageous-deleterious A_1B_1 haplotype. The rest of the population bears the A_0B_0 haplotype. It is assumed that the A_0B_1 haplotype is present at a negligibly small frequency, and while it is not considered in the initial population, it is tracked if it appears by recombination.

A new generation is formed by selecting two parents with probability proportional to their fitness. Recombination between the two parental loci then occurs with Poisson probability r (a Poisson distribution is used to improve the speed of simulations). This is repeated until N new offspring are created. A new generation is created in this way until the A_1B_1 genotype is either fixed or is lost from the population. This entire process is repeated 20,000 times to build up an overall probability of fixation, along with 95% confidence intervals. I focus attention on the processes where the advantageous allele fixes.

Results are plotted in Figure 4.3. Simulation data match up very well to all three solutions for the probability of hitchhiking $P^*(p)$: semi-deterministic equation 4.5, diffusion equation 4.24, and the approximation to the diffusion equation 4.12. All three solutions offer similar results when I changed the population size, as long as ρ, S_d, S_a are held constant. Differences between the solutions only become apparent when selection becomes weak. Stochastic effects then play more of a role, especially where the A_1B_1 haplotype is over-sampled and rises to fixation faster than expected, so that the diffusion with killing 4.24 provides a slightly more accurate solution. Additional figures presented in section 3 of the Supplementary Materials show that the analytical solutions perform less well as selection strengthens in very small populations (e.g., $s_d = 0.1$ with N = 100 or 1000); in these cases, the diffusion approximation assuming weak selection breaks down and the fixation probability of the deleterious allele is underestimated.

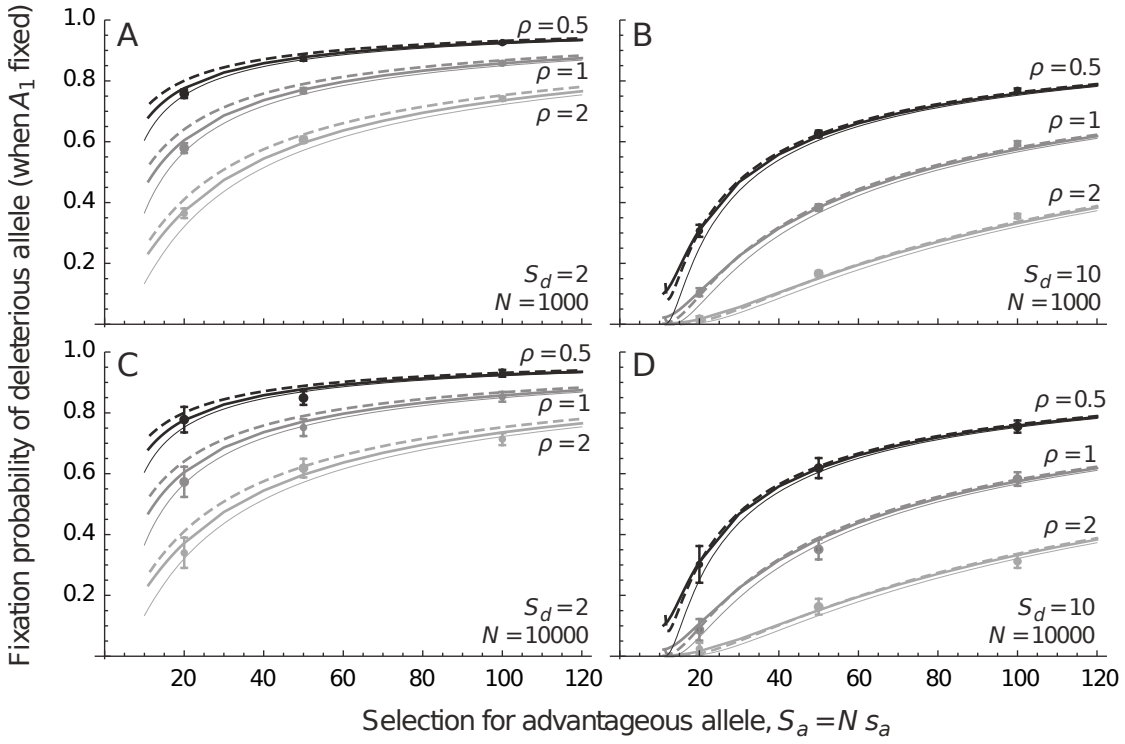


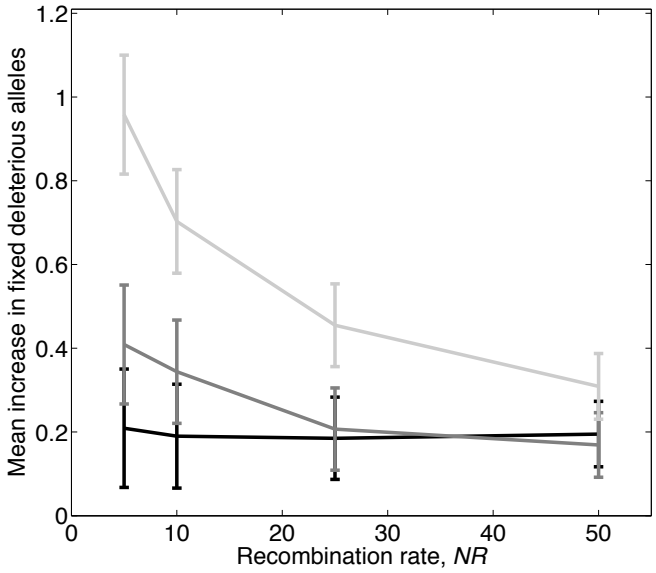
Figure 4.3: Fixation probability of the deleterious B_1 allele, given that the A_1B_1 haplotype appears initially at frequency of $1/N$ and that the A_1 allele is not lost when rare, for different recombination rates $\rho = Nr$. Plots compare the solution to the semi-deterministic model (4.5) (thin solid), the full solution to the diffusion (4.24) (thick solid), the approximation to the diffusion (4.12) (thick dashed), and simulation results based on the Wright-Fisher model (points). Bars indicate 95% confidence intervals here and throughout. Parameters are $N = 1000$ (A and B) and $N = 10,000$ (C and D), with $S_d = 2$ (A and C) and $S_d = 10$ (B and D).

4.5 Multilocus simulations

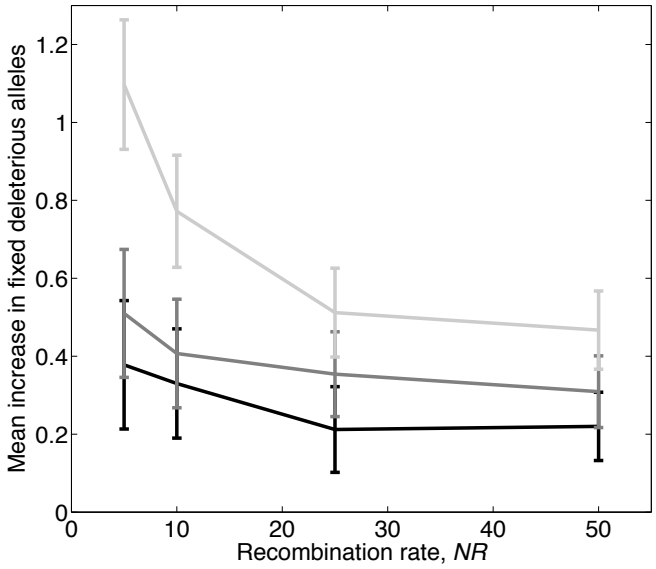
Although the above two-locus models offer tractable results, novel advantageous alleles may arise in genomes with multiple mutant alleles. Therefore, I switch to using multilocus computer simulations to investigate the mutation load generated by the rise to fixation of an advantageous allele, given that such mutations arise at rate U in a genome with total map length R, where each new deleterious mutation is assigned a random position between 0 and R. The methods used for these simulations are based on those in Chapter 3 and detailed in Appendix 4.D.

I then determined the mean number of deleterious alleles that fix along with each beneficial mutation, assuming multiplicative selection. Simulations with different S_a values are compared to the control case, $S_a = 0$, in Figure 4.4 (and 4.5 in the supplementary figures). These results corroborate the two-locus model; the mean number of deleterious mutants that fix declines with the rate of recombination and rises with the strength of selection on the advantageous mutant, S_a. The mean number of fixed deleterious alleles also stays approximately the same as N increases, if the compound parameters S_a, S_d, NR, and NU are held constant.

Increasing the recombination rate also raises the fixation probability of the advantageous mutant (Figure 4.6 in the supplementary figures), which is a well-known result (PECK 1994; BARTON 1995b). Thus recombination is doubly advantageous, as it reduces the number of deleterious alleles that fix in a population following a selective sweep and it increases the likelihood that such an advantageous mutant can establish when rare. This is the likeliest cause of strong selection acting on a modifier for increased recombination in the presence of advantageous and deleterious mutations (Chapter 3).



(a)



(b)

Figure 4.4: The increase in the number of deleterious alleles that fix genome-wide for a given S_a, subtracting off the number that fix in the $S_a = 0$ case, as a function of the total map length NR (see Figure S.2 for the raw data). Only cases where the advantageous allele has fixed are considered. $S_a = 20$ (black line), 40 (dark gray line), or 80 (light gray line). $S_d = 10$, $NU = 50$, and (a) $N = 500$ or (b) $N = 1000$.

4.5.1 Applying results to human genetic data

How likely is deleterious hitch-hiking to occur in nature? To answer this I use human data as an example. Deleterious mutants are maintained at a mutation-selection balance frequency of $q = \mu/s_d$ (HALDANE 1927; WRIGHT 1931), where s_d measures selection against the deleterious allele in heterozygotes. Thus an estimate for the number of deleterious mutants segregating throughout a genome is U/s_d, for U the diploid per-genome deleterious mutation rate, which has been recently estimated as $U = 4.2$ (EÖRY *et al.* 2010).

U measures deleterious mutations arising across the entire genome, with the majority appearing in non-coding regions (EÖRY *et al.* 2010). Thus I assume all deleterious mutations have a fixed, weak value of s_d. This will slightly overestimate the number of deleterious mutants segregating, as I do not consider stronger deleterious mutations that can arise in coding regions (EYRE-WALKER *et al.* 2006; BOYKO *et al.* 2008).

A deleterious allele must have $N_e s_d \geq 1$ in order for selection to overcome the effects of genetic drift (KIMURA 1983). Therefore, assuming deleterious alleles are very weakly selected for, with $N_e s_d = 1$ and human N_e = 10,000 (JORDE *et al.* 1998), I expect $U/s_d = 4.2/0.0001 = 42,000$ such deleterious alleles segregating at any time, roughly half of which lie in each haploid set of 3Gb in the human genome. Note that this value is an overestimate since it is clear that not all deleterious alleles will be weakly-selected for. Including the site of the beneficial mutation, the average distance between two selected sites is thus 142.9kb. Assuming that selected sites are randomly distributed across the genome (i.e., ignoring clustering), this distance would be approximately exponentially distributed. In this case, the closest of the deleterious alleles lying to either side of the beneficial allele would also be exponentially distributed with mean 71.4kb. As a rough guide, the average recombination rate is 1 cM/Mb in a human genome (BROMAN *et al.* 1998), thus the closest deleterious allele lies, on average, at a distance of $N_e r = 7.14$. The fixation probability of the deleterious allele with the

advantageous mutant would then be 18.8% for $N_e s_a = 5$, 37.1% for $N_e s_a = 25$, and 62.1% for $N_e s_a = 100$, obtained by integrating the hitchhiking probability (4.24) over an exponentially distributed distance with mean $N_e r = 7.14$. These calculations are explained in more detail in section 5 of the Supplementary Material. If I assumed $N_e s_d = 10$, then by following similar logic I calculate that the mean distance to the nearest deleterious allele is $N_e r = 71.4$, and the estimated fixation probability of a deleterious allele is 0.8% for $N_e s_a = 25$ and 2.5% for $N_e s_a = 100$.

Overall, these calculations suggest that in humans, deleterious mutants will hitchhike at appreciable frequencies only if they are very weakly selected ($N_e s_d < 10$). However, this is only an initial calculation that deserves to be revised to take into account fine-scale recombination rates (McVEAN *et al.* 2004), the selection strength of mutants being chosen from a distribution of fitness effects EYRE-WALKER and KEIGHTLEY (2007), and clustering of mutations around coding regions. For now, I note that if clustering causes the average recombination distance to a deleterious allele to drop ten-fold, then the hitchhiking probabilities calculated above increase substantially, rising for $N_e s_d = 1$ to 68%, 85%, 94% with $N_e s_a = 5$, 25, 100, respectively, and for $N_e s_d = 10$ to 7%, 20% with $N_e s_a = 25$ and 100. However, note that this calculation does not take into account clustering of deleterious alleles as well.

4.6 Discussion

As long as genetic variance in fitness is present within a population, new beneficial alleles can arise in genomes that, by chance, carry deleterious alleles at linked sites. Consequently, if they remain associated, deleterious alleles can hitchhike to fixation as an advantageous allele sweeps through the population. Even if recombination occurs between the two loci, there can still be a good chance of both alleles fixing, if either the recombinant fails to appear in time or is lost by chance when it does appear.

WILLIAMSON *et al.* (2007) found possible evidence of such hitchhiking causing the high prevalence of the hereditary hemochromatosis mutation C282Y, due to a selective sweep occurring 150kb away from the HFE gene where the deleterious C282Y allele is located.

To my knowledge, this chapter produces the first theoretical study on how recombination affects the hitchhiking to fixation of deleterious alleles. Using both a semi-deterministic and a diffusion approach, I show that, in regions of low recombination, there is a high probability that a deleterious mutant would be swept to fixation if linked to an advantageous mutant (Figure 4.1). This probability approaches one as the deleterious effect s_d tends towards zero and the overall advantage of the A_1B_1 haplotype s_{net} is larger. Outside this parameter range, I find that hitchhiking is likely (greater then 50% chance) if $r \lesssim s_{net}/(4Ns_d)$ (more precisely, equation 4.13). A promising empirical approach would be to investigate areas around the genome that show high d_N/d_S values. It has been postulated that such regions could be subject to recurrent sweeps (NIELSEN 2005). If deleterious alleles do hitchhike, then around these sites there should be signs of increased load, such as lower frequency of optimal codon usage. Such a negative relationship between dN and optimal codon usage was found in *Drosophila* by BETANCOURT and PRESGRAVES (2002), although other mechanisms could explain this finding, such as an increased frequency of weakly-selected deleterious mutations being introduced by recurrent mutation.

Furthermore, we determined that the hitchhiking of tightly linked deleterious alleles reduces the region in which the sweep is likely to fix surrounding sites (compare equation 4.15 to equation 4.13). This is important, as it implies that deleterious hitchhiking can alter experimental estimates of the strength of such sweeps. A potential example of these effects was reported by CLEGG *et al.* (1980), who found that linkage disequilibrium in *Drosophila melanogaster* broke down more quickly than expected (geometric decay at a ratio $1 - r$), based on the surrounding markers being neutral and on meas-

ured recombination rates between the selected and neutral markers. This observation could be explained by recombination untangling advantageous alleles from deleterious backgrounds (see also Figure 4.8). Further work is warranted to explore the impact of neighbouring selected sites on patterns of neutral sequence variability in a fully multi-locus framework. In particular, a full treatment requires an exploration not only of the primary effects of a selective sweep at a focal site, but also of how hitchhiking of deleterious alleles followed by recombination can cause secondary sweeps as wildtype alleles reestablish themselves at surrounding sites.

My work also sheds light on the results found in Chapter 3, showing that a modifier gene for increased recombination is more likely to fix in a population that is subject to both deleterious and advantageous mutation, compared to the deleterious-only mutation case (KEIGHTLEY and OTTO 2006). The increased selection acting on a recombination modifier when both deleterious and advantageous mutants are present together, compared to when just deleterious or just advantageous mutations are present, suggests that uncoupling advantageous mutants from deleterious backgrounds provides a substantial amount of selection on a recombination modifier (PECK 1994; and Chapter 3).

These preliminary calculations suggest that in obligately sexual species with long genetic map lengths (such as the human genome), recombination is frequent enough to prevent all but weakly deleterious mutants from hitchhiking with advantageous mutants. These calculations assumed, however, that mutations affecting fitness arise at equal rates throughout the genome, which ignores the clustering of selectively-constrained sites near genic regions. If recombination rates between selected sites are low, either because of this clustering or because of coldspots in recombination, the probability that deleterious alleles hitchhike to fixation rises substantially. Similarly, in species that frequently inbreed (e.g., selfing) or reproduce asexually, the effective amount of recombination may be much lower, substantially increasing the probability of deleterious alleles hitchhiking to fixation. In asexuals with no recombination, the subsequent mutation

accumulation can be extremely detrimental (HADANY and FELDMAN 2005).

In conclusion, sex and recombination both enhance the probability of beneficial alleles establishing and hinder the fixation of deleterious alleles within a lineage. If this can be shown empirically to be a potent selective force on recombination rates, then this would provide key insight into why sex and recombination are prevalent, which remains an open question in evolutionary genetics (OTTO 2009).

4.A Derivation of $\Pi(t)$, the probability of establishment of a recombinant haplotype

When the recombinant A_1B_0 haplotype is produced, it appears within a population that is already changing due to the spread of the A_1B_1 haplotype. Thus, I cannot calculate the probability of fixation of the recombinant A_1B_0 haplotype based solely on its fitness $1 + s_a$ relative to the current population mean $1 + p(t)(s_a - s_d)$. Rather, I must also account for future changes in the population mean fitness as the A_1B_1 haplotype rises in frequency. To do so, I develop a time-inhomogeneous branching process that explicitly follows the dynamics of $p(t)$ (given by equation 4.1) that occur after the appearance of the recombinant A_1B_0 haplotype. A previous diffusion analysis by KIMURA and OHTA (1970) also calculated the fixation probability for a favorable allele whose benefit declined over time, but the focus of their analysis was on a case where selection declines linearly over time, whereas here, the selection coefficient favoring A_1B_0 declines according to a logistic function of time, given by $s(t) = s_a - p(t)(s_a - s_d) + O(s^2)$ if assuming weak selection.

Let $\Pi(t)$ be the fixation probability of the recombinant A_1B_0 haplotype at generation t, given that the current frequency of the A_1B_1 haplotype is $p(t)$. In a population of constant size, the average parent has one surviving offspring, but I assume that the A_1B_0 haplotype is more fit and so has an average of $1 + s(t)$ offspring. Using branching

process logic (HALDANE 1927), the recombinant A_1B_0 haplotype will ultimately be lost (with probability $1 - \Pi(t)$) if and only if all j offspring inheriting the haplotype also fail to leave any descendants over the long run (with probability $(1 - \Pi(t+1))^j$). Assuming a Poisson distribution for the number of offspring j and summing over this distribution, I obtain a recursion for $\Pi(t)$:

$$\begin{aligned} 1 - \Pi(t) &= \sum_{j=0}^{\infty} e^{-(1+s(t))} \frac{(1+s(t))^j}{j!} (1 - \Pi(t+1))^j \\ &= exp[-(1+s(t))\ \Pi(t+1)] \end{aligned} \tag{4.16}$$

Solving for $\Pi(t+1)$ and subtracting $\Pi(t)$, I obtain the change in fixation probability over time, which I assume is slow enough that it can be well approximated by the differential equation:

$$\frac{d\Pi}{dt} = -\frac{ln[1 - \Pi(t)]}{1 + s(t)} - \Pi(t) \tag{4.17}$$

With weak selection ($s(t) \ll 1$), $\Pi(t)$ is of the same order as $s(t)$ and the above simplifies to

$$\frac{d\Pi}{dt} = -\frac{1}{2}\Pi(t)^2 + s(t)\ \Pi(t) + O(s^2) \tag{4.18}$$

(BARTON 1995b). This differential equation can be solved when selection on the recombinant haplotype varies according to $s(t) = s_a - p(t)(s_a - s_d)$ by first replacing the variable t with the variable p, using the chain rule, and $dp/dt = s_{net}p(1-p)$ (Supplementary Materials, section 1). To leading order in the selection coefficients (i.e. discarding terms of order s^2 or raised to a higher power), the resulting solution for the fixation probability of the recombinant A_1B_0 haplotype is given by equation 4.2.

4.B Deriving the diffusion process with killing conditional on fixation of the A_1 allele

Conditioning on the fixation of A_1 implies that either the A_1B_1 haplotype fixes (if the process is not killed) or the recombinant successfully establishes itself and leads to the fixation of the A_1B_0 haplotype (if the process is killed). Either way, the A_1B_1 haplotype cannot be lost while it is rare. I must therefore adjust the drift term in the diffusion, $M(p)$, to account for the fact that the A_1B_1 haplotype will, on average, rise more rapidly when rare amongst those processes where the A_1B_1 haplotype is not lost. The variance term $V(p)$ and the killing term $K(p)$ are unchanged in the conditioned model, as these terms depend only on the current frequency of the A_1B_1 haplotype and not on its ultimate fate. From equation 9.5 in chapter 15 of KARLIN and TAYLOR (1981), the conditional drift term $M^*(p)$ is given by

$$M^*(p) = S_{net}\, p(1-p) + \frac{s(p)}{S(p)}\, p(1-p) \tag{4.19}$$

where

$$s(p) = exp\left[-\int_0^p \frac{2M(\eta)}{V(\eta)} d\eta\right] \tag{4.20}$$

$$S(p) = \int_0^p s(\xi)\, d\xi \tag{4.21}$$

Here, the values of $M(p)$ and $V(p)$ are for the unconditional diffusion process as outlined in the main part of the chapter. Plugging these terms into equations 4.20 and 4.21 and evaluating the integrals, I obtain the conditional drift term:

$$M^*(p) = S_{net}\, p(1-p)\, \frac{1+e^{-2pS_{net}}}{1-e^{-2pS_{net}}} \tag{4.22}$$

This revised drift term is then placed in equation (4.8), along with the variance and killing terms, which remain unchanged. Dividing the result by $p(1-p)$ yields equation 4.11 in the main text.

The conditional diffusion process requires some care, however, with the boundary conditions. The probability that the process is not killed given that the A_1B_1 haplotype is fixed remains one, $P^*(1) = 1$, as before. Conditioning assumes, however, that the $p = 0$ boundary is never reached. Rather than assigning $P^*(0)$, I instead assume that $P^*(p)$ varies little over very small values of p, given that the process will ultimately reach $p = 1$ if it is not killed. Thus, I use $dP^*(0)/dp = 0$ as a second boundary condition.

Solving equation 4.10, I find that the probability that the process is never killed, regardless of whether A_0 or A_1 ultimately fixes is:

$$\begin{aligned} P(p) = \big(&U^0_{-\omega}[-2(pS_{net}+S_d)]\ (L^{-1}_{\omega}[-2S_d] - L^{-1}_{\omega}[-2(S_{net}+S_d)]) \\ &-L^{-1}_{\omega}[-2(pS_{net}+S_d)](U^0_{-\omega}[-2S_d] - U^0_{-\omega}[-2(S_{net}+S_d)])\big)\ / \\ &\big(U^0_{-\omega}[-2(S_{net}+S_d)]\ L^{-1}_{\omega}[-2S_d] - U^0_{-\omega}[-2S_d]\ L^{-1}_{\omega}[-2(S_{net}+S_d)]\big)\,, \end{aligned} \quad (4.23)$$

while the solution to equation 4.11, conditioned on the fixation of the beneficial A_1 allele, is:

$$\begin{aligned} P^*(p) = &\left(\frac{1-e^{-2S_{net}}}{1-e^{-2pS_{net}}}\right) \times \\ &\frac{U^0_{-\omega}[-2(pS_{net}+S_d)]\ L^{-1}_{\omega}[-2S_d] - U^0_{-\omega}[-2S_d]\ L^{-1}_{\omega}[-2(pS_{net}+S_d)]}{U^0_{-\omega}[-2(S_{net}+S_d)]\ L^{-1}_{\omega}[-2S_d] - U^0_{-\omega}[-2S_d]\ L^{-1}_{\omega}[-2(S_{net}+S_d)]} \end{aligned} \quad (4.24)$$

Here, $U^b_a[z] = U[a,b,z]$ is the Tricomi confluent hypergeometric function, $L^{\alpha}_n[x]$ the generalised Laguerre polynomial (ABRAMOWITZ and STEGUN 1970), and ω is the

compound parameter given by equation 4.6 in the main text. Additional details regarding the derivation and solutions for these equations are provided in a *Mathematica* 6.0 file (Supplementary Materials, section 2).

4.C Testing the effect of recombination on the fixation of a linked, neutral allele

I investigate the effect of deleterious hitchhiking on neutral sequence, and how recombination affects this, by extending the two-locus simulations. In these new simulations, there are three linked loci; the deleterious site, a neutral locus and a third locus where the sweep is present, all separated by a recombination distance r. Additionally, there is now recurrent mutation occurring at rate μ at the deleterious locus, with no back-mutation.

The initial set-up is different as well: initially the deleterious allele is present at a fixed mutation-selection balance frequency of μ/s_d (WRIGHT 1931), or 50% if $s_d = 0$. When the advantageous allele is introduced in a single copy, it is placed within a random individual that does not necessarily carry a deleterious allele. This is because I want to measure the difference in diversity due to background selection, averaged over all possible initial backgrounds. A neutral allele is also introduced in the individual in which the sweep first arises. This neutral marker allows us to measure the extent to which the initial neutral diversity is reduced to the single allele that happens to be adjacent to the new mutation.

The population then undergoes the same cycle of selection, recombination, then mutation at the deleterious locus. The sweep is tracked until it is fixed or lost. If fixed, the frequency of the neutral allele is noted. The sweep is then reintroduced 1,000,000 times, and the mean final frequency of the neutral allele is measured.

If the mean frequency is near one, this implies that little recombination has taken place between the neutral and selectively favoured allele over the course of the sweep.

A lower value implies that a higher level of recombination has taken place, resulting in the advantageous allele becoming separated from the neutral allele that it was originally linked to. The results of the main text predict that recombination must be even tighter (lower ρ_{crit}, see Figure 4.2) for hitchhiking to fix nearby deleterious alleles, compared to the case of nearby neutral alleles, suggesting that those recombination events that do occur in the presence of surrounding selected sites are more likely to establish themselves, on average, within the population during a selective sweep. The increased establishment of recombinant chromosomes should result in reduced effects on linked neutral diversity as well. In particular, in these simulations, I predict that a beneficial allele will not drag a neighboring neutral allele to as high a frequency in the presence of another selected site in the surrounding region.

Figure 4.8 plots the results of these simulations. In all cases tested, I verify the prediction that if $S_d > 0$, the initially linked neutral allele does not sweep to as high a frequency on average, compared to the $S_d = 0$ case. This indicates that the diversity present at linked sites is more likely to be preserved by recombination when beneficial alleles arise at sites surrounded by others subject to selection. The reductions observed in these simulations are modest; this is due to there being only one linked deleterious site, with a low mutation rate ($\mu = 0.0005$). A larger effect would be observed if the mutation rate was higher or more linked deleterious loci were present.

In summary, I argue that linked selected sites should reduce the impact of a selective sweep has on surrounding neutral diversity. My reasoning focuses on the increased probability that recombinant chromosomes will establish within the population during the sweep, because they uncouple a beneficial allele from any deleterious alleles within its genetic background (main text, BARTON (1995b)). Verifying that this is indeed the case in a multi-locus framework deserves future work. In particular, these simulations have not taken into account the variation in fitness among recombinant chromosomes due to additional selected sites throughout the genome, beyond the one neighboring se-

lected site. Furthermore, I have not taken into account the cascade of secondary sweeps that occur whenever fitter recombinants arise and drag along with them their own suite of alleles.

4.D Methods used for Multilocus Simulations

Initially there is a haploid population of N chromosomes with an infinite number of loci per chromosome. Each locus has a wildtype allele or a deleterious allele with selection s_d acting against it. Fitness is multiplicative, so initially in the absence of the advantageous allele the fitness of an individual is $(1 - s_d)^k$, where k is the number of deleterious mutants present in an individual chromosome.

New generations are created by selection, recombination, then mutation. Two parents are chosen with replacement from the population, with probability proportional to their fitness. Recombination then occurs, with the number of crossovers across the chromosome selected from a Poisson distribution with mean R (where R is the total genome map length). One of these parents is selected to be the template for the offspring genome. Each mutant has a map position assigned to it as it appears, which is drawn from a uniform $[0, 1]$ distribution. For each crossover event, the position of the recombination event is also drawn from a $[0, 1]$ uniform distribution. The allelic states are then swapped at sites whose map distances exceed the recombination distance. If two crossovers are chosen, locus states are swapped at sites whose map distances lie between the two recombination distances. The number of crossovers is capped at two for ease of computing, which leads to little loss of accuracy if R is small (Chapter 3).

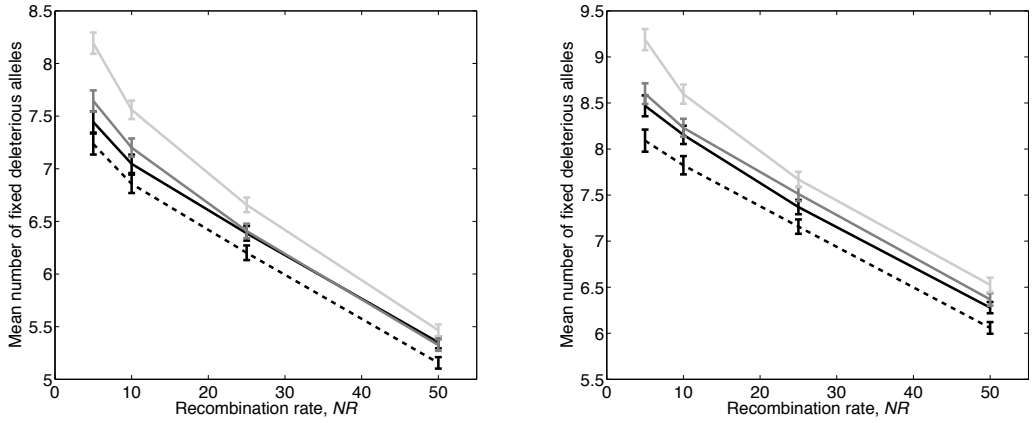
For each offspring, the number of new deleterious mutants is chosen from a Poisson distribution with mean U. Each new mutant is assigned to a new locus. Back mutation also occurs at a deleterious allele with probability $\mu = 10^{-8}$, which is approximately equal to the per-locus mutation rate in humans (EÖRY *et al.* 2010). Overall, the whole

cycle is repeated N times to repopulate the gene pool.

There is an initial burn-in of $2N$ generations, so that the population reaches a mutational steady-state. A 'garbage collection' routine is executed every 50 generations during the burn-in; mutants that are lost from the population are cleared to free memory, as well as deleterious mutants that have fixed, so that I do not consider deleterious mutants that accumulate through Muller's Ratchet before the advantageous allele is introduced. Following the burn-in, the state of the population is saved and mutation is turned off, so deleterious mutants that do fix tend to be driven to fixation by hitchhiking, rather than an on-going ratchet mechanism. An advantageous allele is then added to a random chromosome at a random site. This allele increases the fitness of the host chromosome to $(1 + s_a)(1 - s_d)^k$.

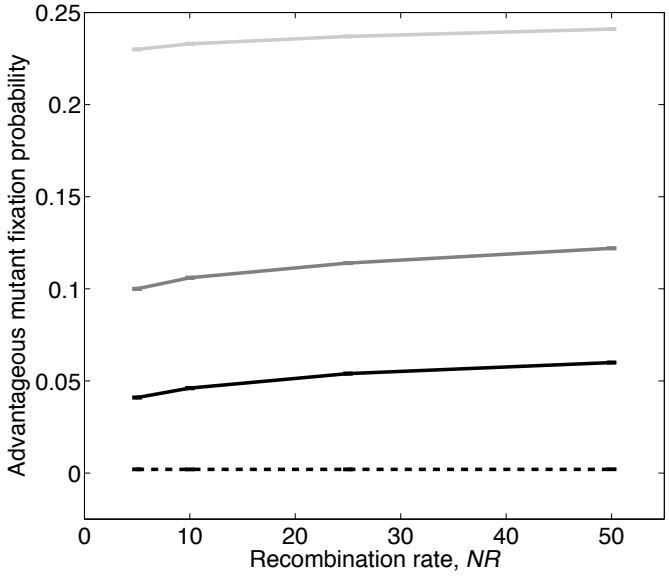
The advantageous allele is then tracked until it is fixed or lost from the population. During this time, a different garbage collection routine is run every 50 generations, which only clears mutants lost from the population in order to free memory. Fixed deleterious alleles are not cleared at this stage, so the mean number that fix with the sweep can be measured. If the advantageous mutant reaches fixation, then all remaining deleterious mutants are tracked until they are fixed or lost, to determine how many deleterious mutants fix. The advantageous mutant is reintroduced from the burn-in population 3,000 times, and its fixation probability is calculated, along with the average number of fixed deleterious mutants. This is repeated for 4,000 burn-ins to build a probability distribution for these statistics.

4.E Supplementary figures

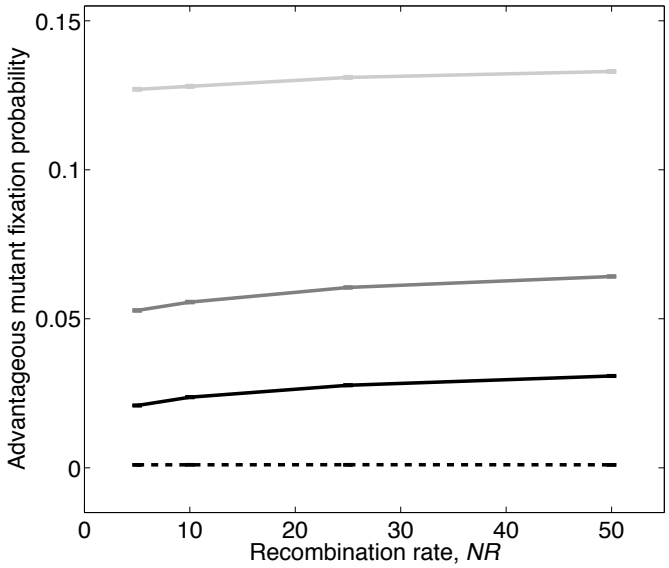


(a) (b)

Figure 4.5: The mean number of deleterious alleles that fix genome-wide following the completion of a successful selective sweep, given as a function of the recombination rate NR (see Figure 4.4 for data presented relative to $S_a = 0$). $S_a = 0$ (black dashed line), 20 (black solid line), 40 (dark gray) or 80 (light gray). $S_d = 10$, $NU = 50$, and (a) $N = 500$ or (b) $N = 1000$.



(a)



(b)

Figure 4.6: Fixation probability of the advantageous mutant in multilocus simulations, as a function of the recombination rate NR. $S_a = 0$ (black dashed line) 20 (black solid line), 40 (dark gray) or 80 (light gray). $S_d = 10$, $NU = 50$, and (a) $N = 500$ or (b) $N = 1000$.

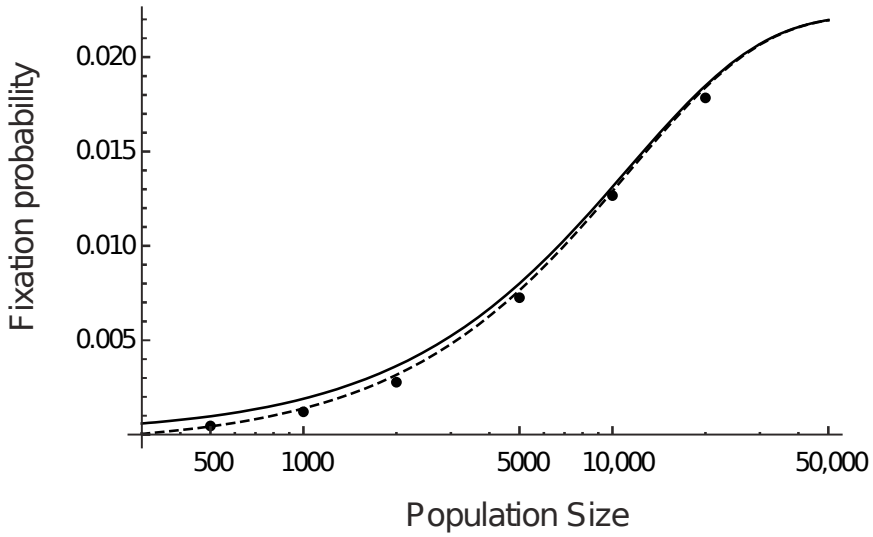


Figure 4.7: Fixation probability of a recombinant carrying two beneficial mutations. Using the notation of YU and ETHERIDGE (2010), $s\gamma$ is the selective advantage of the first beneficial mutation to occur, s is the selective advantage of the second beneficial allele, and $s(1+\gamma)$ is the selective advantage of the recombinant double mutant. The semi-deterministic solution 4.5 (solid curve) and the diffusion solution 4.12 (dashed curve) are presented alongside simulation results (dots) for the parameters considered in Figure 4a of YU and ETHERIDGE (2010), where $s_{net} = s(1-\gamma)$ equals the difference in fitness between 10 and 01 individuals, and where $s_a = s$ and $s_d = s\gamma$ give the selective advantage of the double mutant when it appears in a population predominantly composed of 10 and 01 individuals, respectively. These curves are multiplied by the establishment probability of the second beneficial allele, given by equation 4.2 with the selection coefficients now reflecting the advantage of 10 spreading within a population of 00 individuals ($s_{net} = s\gamma$), within which a 01 mutant appears with advantage $s_a = s$ over the 00 wildtype. Parameters as in Figure 4a of YU and ETHERIDGE (2010): $s = 0.02$, $\gamma = 0.8$, $r = 0.00001$, with a starting frequency of haplotype 10 of 0.2 at the time that the second beneficial mutation appears in a 01 haplotype.

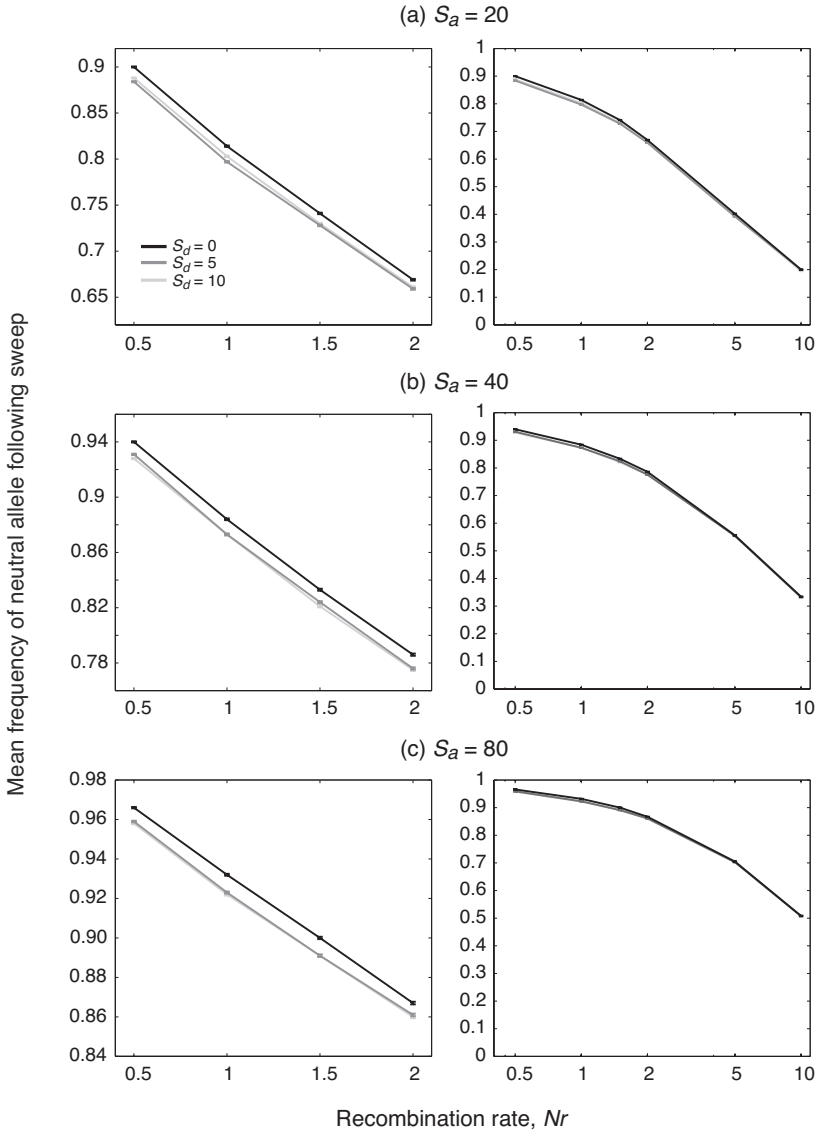


Figure 4.8: The mean frequency of a linked neutral allele following a successful selective sweep, given as a function of the recombination rate Nr between different sites. The left- and right-hand panels report the same simulations, with the left-hand panels zoomed into the region (tighter linkage) where the impact of a neighboring selected site on the patterns of neutral diversity was greatest. $S_d = 0$ (black line), 5 (dark gray line) or 10 (light gray line). $N = 1000$, $\mu = 0.0005$, with (a) $S_a = 20$, (b) $S_a = 40$, or (c) $S_a = 80$.

Chapter 5

Can weak population structure protect sexual populations from asexual invasion?

Chapter abstract

Although there is no know general explanation as to why sexual populations resist asexual invasion, previous work has shown that sexuals can outcompete asexuals in structured populations. However, it is currently unknown whether sexuals, suffering up to a twofold fecundity cost, can be maintained in populations with weak structure that is commonly observed in nature. Here, I investigate the conditions under which obligate sexuals resist asexual invasion in structured populations subject to recurrent deleterious and advantageous mutations. I determine the population structure needed to disfavour asexuals, as calculated using the average F_{st} between all pairs of demes. I show that levels of F_{st} needed to maintain sex decrease as the population size increases, but these levels are generally very high compared to those observed in field studies, although they could be considerably lower in very large populations. Lower F_{st} values are needed to maintain sex if demes are arranged over one dimension, than if spread out over a two-dimensional torus. However, if a small proportion of migrants are able to move to any deme, the critical level of population subdivision appears to be similar for the two types of structure considered. Asexual fixation probability drops sharply with higher deleterious mutation rates, since this increases the mutation load. Asexuals are most strongly selected against if deleterious mutants have intermediate selective strengths, which maximises the effect of Muller's Ratchet.

5.1 Introduction

Explaining the evolution and ubiquity of sex has been one of the most difficult problems in evolutionary biology. Sexuals suffer a variety of costs, including a twofold cost as a consequence of only transmitting half of their genes to offspring. This can result in a reduced fecundity in a sexual population (MAYNARD SMITH 1978), as well as a 50% reduction in the relatedness between parents and offspring in anisogamous populations (WILLIAMS 1975; LIVELY and LLOYD 1990). Asexual females, on the other hand, pass on their entire genome during reproduction, so should quickly outcompete sexuals, all else being equal. Although the twofold cost can be partially compensated by paternal resource contributions, a convincing theory for the evolution of sex must demonstrate that costs of this magnitude can be overcome.

Of the many theories that have been proposed to explain the maintenance of costly sex, three have received the most attention on the grounds of their generality and potential explanatory power. First, costly sex can be maintained if the genomic rate of deleterious mutation (U) is high enough, and if there is synergistic epistasis between deleterious mutations (KONDRASHOV 1982). However, the current available evidence suggests that epistatic interactions do not appear to be widespread in nature (KOUYOS *et al.* 2007). Secondly, sex could have evolved as a defense against parasites (the 'Red Queen' hypothesis), since sex creates new genotypes that can resist infection (HAMILTON *et al.* 1990). Until recently, it was thought that this mechanism could only select for costly sex if selection is moderate to strong in hosts and parasites (MAY and ANDERSON 1983; OTTO and NUISMER 2004), and if parasites rapidly adapt to infect those with novel genotypes (BARTON 1995a). However, recent theoretical work has shown that parasitic infection can maintain sex under a wider range of realistic scenarios, such as if selection against noninfecting parasites is strong (SALATHÉ *et al.* 2008b), and if density-dependent virulence is assumed (LIVELY 2009, 2010b). There also exists a large body of empirical and experimental evidence demonstrating that sex evolves

in the presence of parasitic infection (JOKELA *et al.* 2009; KING *et al.* 2009; MORRAN *et al.* 2009, 2011). Thirdly, sex and recombination break down negative genetic associations between linked loci, generated by selection and drift in a finite population (Hill-Robertson effects) (HILL and ROBERTSON 1966). By reducing these associations, modifier genes that increase recombination can increase the effectiveness of selection and spread throughout a population (KEIGHTLEY and OTTO 2006; ILES *et al.* 2003; and Chapter 3). Experimental evidence also suggests that breaking down negative associations between linked loci can overcome significant costs of sex (COLEGRAVE 2002; POON and CHAO 2004; GODDARD *et al.* 2005; MORRAN *et al.* 2009).

Although Hill-Robertson effects have gained substantial acceptance as an important driver of the evolution of recombination (OTTO 2009; BARTON 2010), the maintenance of costly sex in the face of invasion by rapidly reproducing asexual mutants remains unsolved. Geographically structured populations are better able to resist asexual invasion (PECK *et al.* 1999; SALATHÉ *et al.* 2006; MARTIN *et al.* 2006), because structure increases the fixation time of asexual lineages, allowing more time for accumulation of deleterious mutations via Muller's Ratchet (FELSENSTEIN 1974; KEIGHTLEY and OTTO 2006). This could therefore be a plausible general explanation for the maintenance of sex, because all populations are subdivided to some extent. However, previous work made no attempt to relate the level of subdivision relative to what is expected in natural populations, which is generally measured using Wright's F_{st} statistic (WRIGHT 1951). Therefore, it remains to be investigated whether costly sex can be maintained with realistic levels of population subdivision, and how the type of population structure affects the mean level of F_{st} needed to maintain sex.

Here, I extend previous simulations that investigated the evolution of a recombination modifier gene (Chapter 3), in order to determine whether models involving realistic levels of population subdivision can maintain sex. Specifically, I model the invasion of an asexual mutant into a subdivided, obligate sexual population. I compare various types

of structure, and examine the additional effect that advantageous mutation have on the maintenance of sex. I also compare the critical level of population subdivision needed to maintain sex, as measured using the average pairwise F_{st} value obtained between each pair of demes. These statistics can be used to inform as to whether realistic levels of population subdivision can maintain sex, in population sizes that I am currently able to simulate.

5.2 Materials and Methods

5.2.1 Basic simulation setup

The simulation was based on previous work that investigated the evolution of a recombination modifier in an asexual population (Chapter 3). I only outline the basic simulation methods in this chapter, and detail the changes made in order to study the maintenance of sex in structured populations. Initially there were N mutant-free haploid chromosomes, each having 100 equally-spaced linked loci. New generations were formed by selection, recombination (if present) and mutation to create N offspring.

As in Chapter 3, mutations were entirely deleterious (as in KEIGHTLEY and OTTO (2006)), entirely advantageous (similar to ILES *et al.* (2003)), or there was a mixture of advantageous and deleterious mutations. In the latter case, the ratio of advantageous to beneficial mutations was equal to $x = k/s_a$ (for $k = 0.00023$ and s_a the selection strength acting on the advantageous allele, as used in HARTFIELD *et al.* (2010)). The function k/s_a reflected the view that strongly advantageous mutants are less likely to appear than weakly selected ones in nature (ANDOLFATTO 2007; JENSEN *et al.* 2008). Thus, the overall proportion of advantageous mutation present was $(x/(1 + x))$. The number of mutants introduced into an individual offspring was drawn from a Poisson distribution, with a mean set to U if there were only deleterious mutations, Ux if there were only advantageous mutations, and $U(1 + x)$ if there was a mixture of deleterious

and advantageous mutations. Each site was equally likely to acquire a new mutation, and the fitness effects of new mutations were multiplicative, with no epistasis.

In most simulations, I assumed U = 1.0, which is comparable with estimates for several multicellular eukaryotes (DENVER *et al.* 2004; HAAG-LIAUTARD *et al.* 2007; EÖRY *et al.* 2010). I either assumed that mutations are deleterious (with selection coefficients s_d = 0.01 in most cases) or that there is a mixture of deleterious and advantageous mutations (the latter with selection coefficients s_a = 0.01 by default), with multiplicative fitness interactions across loci. This value of s_a leads to advantageous mutations comprising a small fraction of all mutations (around 2.2% in most simulations), which lies within the recently-obtained estimates of the proportion of advantageous amino-acid-changing mutations in *Drosophila melanogaster* of 0.5% and 3.5% (SCHNEIDER *et al.* 2011). However, these estimates are likely to be conservative because many adaptive mutations are believed to occur in noncoding DNA (SELLA *et al.* 2009). I also examined the sensitivity of the asexual fixation rate to changes in the mutation rate and strength of selection on deleterious and advantageous mutations.

5.2.2 Invasion of asexuals into a structured sexual population

The population was subdivided into D demes, with $N_D = N/D$ individuals per deme, and an overall migration rate m between demes. 25 demes were used by default, although the number was later varied to determine how the number of demes affects the F_{st} value needed to maintain sex. Demes were either arranged in a one-dimensional circular array, or over a two-dimensional torus (as outlined in Figure 5.1).

Sexual populations first underwent a cycle of selection, recombination and mutation according to standard Wright-Fisher dynamics (FISHER 1930; WRIGHT 1931), but with selection acting over each deme. New offspring were formed by picking one parent with probability proportional to its fitness. If this is an asexual then it was cloned, otherwise a second parent was selected, then outcrossing occurred, with the number of crossov-

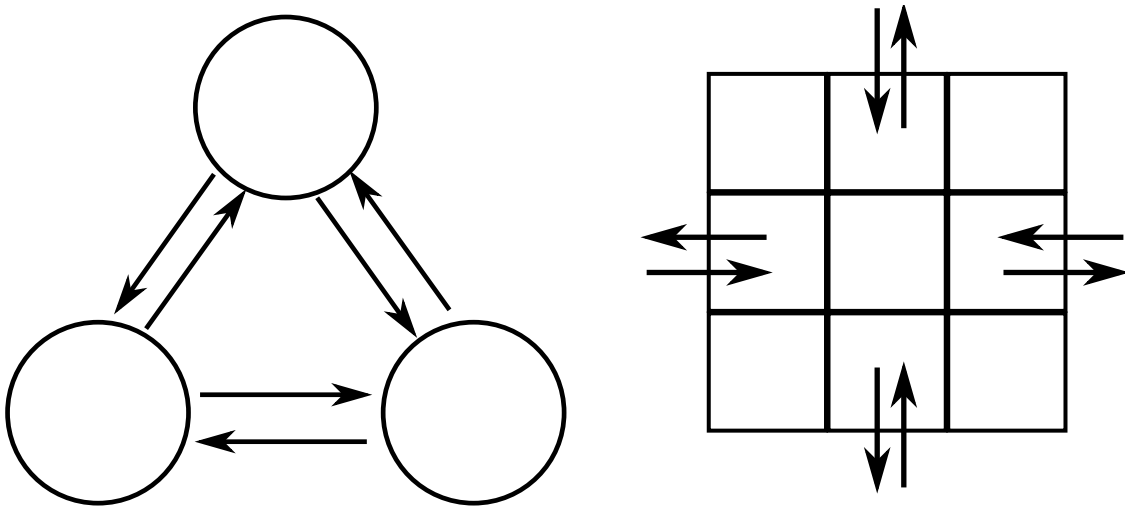


Figure 5.1: Schematic of the two types of population structure investigated. In the one-dimensional model (left), demes are arranged in a circular formation, with migration rate m between demes, and migrants equally likely to migrate to one of two neighbours. In the two-dimensional model (right), demes are arranged on a grid, and migrants can transfer to one of four neighbouring demes.

ers chosen from a Poisson distribution with mean 1. This genomic crossover rate is lower than that observed in higher eukaryotes such as humans (BROMAN *et al.* 1998; MCVEAN *et al.* 2004), suggesting that the observed benefits in breaking down selection interference might be conservative in comparison to natural populations. Mutants were then added as described above. This was repeated N_D times to create a new population within each deme. Sexuals suffered a twofold fitness cost $C = 2$ unless stated otherwise. This cost reduced the probability that a particular parent would be chosen (so that the fitness of a sexual, $w_{sex} = w_{asex}/2$). After reproduction was completed for all demes, migration occurred. The number of migrants was chosen from a Poisson distribution with mean mN_D. By default, each migrant moved to a randomly chosen neighbouring deme, and each neighbour was equally likely to be a migrant destination. If long-distance migration occurred, then each migrant was moved to a randomly-chosen deme with probability p. The lifecycle was repeated for $10N$ generations in order to

create a burn-in population. A relatively long burn-in time was used to ensure that the population's steady-state F_{st} value was reached (WHITLOCK and MCCAULEY 1999) (see Figure 5.7 in the supplementary figures for evidence that such a steady-state was approached).

The state of the population was then saved, then an individual in a single deme was changed into an asexual. This asexual was tracked until it was fixed or lost from the population. The asexual was introduced N times into the burn-in population to establish a fixation probability u, which is reported here as that relative to the fixation probability of a neutral mutant, $u^* = 1/N$. This was repeated for 40 individual burn-ins. Larger populations were run with more burn-ins (50) and fewer reintroductions per burn-in ($0.4N$), to reduce the standard errors reported for the asexual fixation probability.

5.2.3 Measuring F_{st} in simulations

WRIGHT (1951) introduced F_{st} as a measure of the degree of mixing in a structured population (HARTL and CLARK 2007). For example, $F_{st} = 1$ indicates that demes are isolated and $F_{st} = 0$ indicates that they are fully mixed. Appendix 5.A reviews F_{st} values found in studies of natural populations. In groups of animal and fish with little physical distance between them, F_{st} values tend to be quite small (usually less than 0.01). For physically more distantly related animal populations, as well as plant populations that have limited gene flow between them, F_{st} values tend to be larger. In these cases F_{st} frequently lies between 0.1 and 0.5. I use F_{st}, as it is a standard metric for population subdivision that is frequently estimated in empirical studies of natural populations. It can therefore be used to relate the results from my simulations to surveys of natural populations, as well as informing to what extent F_{st} is related to the probability of asexual fixation for different types of population structures. In particular, WHITLOCK (2002, 2003) demonstrated that F_{st} accurately predicts how population structure affects the fixation probability and fixation time of weakly-favourable alleles.

It can therefore reflect how much population structure is needed to delay the spread of an invading asexual so that it eventually becomes disadvantageous.

In order to measure F_{st} within the simulations, new neutral alleles were constantly introduced via mutation; the number added is drawn from a Poisson distribution with mean $2/N$, so that on average two new mutations were added per generation. This arbitrary value enabled an adequate number of neutral alleles to be sampled in order to obtain an accurate estimate of F_{st}. Each new mutant was assigned to its own unique biallelic locus, with a map distance drawn from a uniform $[0, 100]$ distribution (so that this locus also lay within one of the 100 loci at which selected mutations accumulated). Neutral mutants that had either become fixed or lost from the population were cleared every 10 generations.

Every $N/4$ generations for the first $8N$ generations of the burn-in, and every $N/20$ generations for the remaining $2N$ generations, the pairwise F_{st} between each possible pairs of demes in the population was measured. I use the $\hat{\theta}$ estimator to calculate F_{st} (WEIR and COCKERHAM 1984); for each neutral locus k, I measured the sample variance s_k^2 in neutral allele frequency between a pair of demes, and the mean allele frequency $\overline{p}_k$. $\hat{\theta}$ is then calculated by summing the relevant numerators and denominators over all neutral loci, then taking the quotient as shown in Equation 5.1:

$$\hat{\theta} = \frac{\sum_k s_k^2}{\sum_k [\overline{p}_k(1-\overline{p}_k) + s_k^2/2]} \tag{5.1}$$

Note the presence of the $s_k^2/2$ term in the denominator, which is needed to correct for the fact that F_{st} is sampled over only two demes. If measuring $\hat{\theta}$ over a large number of demes, this term would go to zero and the formula would equate to the standard F_{st} calculation of $\sigma_p^2/(\overline{p}(1-\overline{p}))$, for σ_p^2 the population variance in allele frequency (HARTL and CLARK 2007). I sum numerators and denominators separately for each locus then divide by the sums to correct for cases where a neutral allele has fixed in both demes (WEIR and COCKERHAM 1984).

To measure the mean average pairwise F_{st} for a particular migration rate, for each burn-in population the F_{st} values recorded for the final $2N$ generations were averaged to produce the F_{st} estimate for that run. These are then averaged over all burn-in populations to obtain a final mean estimate.

5.2.4 Finding the critical F_{st} value that makes sex advantageous

I measured the average pairwise F_{st} for each burn-in population. By plotting these parameters as a function of time, I can check whether they reached a steady-state. Figure 5.7 in the supplementary figures shows typical plots of the average and maximum F_{st} value for a small population ($N = 1000$) and a large population ($N = 20,000$). I observed that F_{st} measurements increased from an initially low value to approach a steady-state, but this took a large proportion of the $10N$ generations if the population was small (Figure 5.7(a)). If the population was large, then a steady-state was arrived at earlier, with subsequent measurements fluctuating around a mean value (Figure 5.7(b)).

By examining the population at regular intervals, I measured the F_{st} values for different migration rates, and plotted the asexual fixation probability u/u^* as a function of them. Figure 5.8 in the supplementary figures shows such a plot for $N = 10,000$, for a case where demes are spread out over one dimension and individuals are subject to advantageous and deleterious mutation. By fitting an exponential curve to the data I obtained the critical average F_{st} value for which asexuals became selected against; that is, the F_{st} value where the fixation probability of an asexual mutant equalled that for a neutral allele ($u/u^* = 1$). For example, in Figure 5.8 the critical value is $\sim$0.59. Each data set was bootstrapped 1000 times to obtain estimates of the standard error for each critical F_{st} value.

5.3 Results

5.3.1 Exploring the critical F_{st} needed to maintain sex changes with population size

In the absence of population structure, I found that for populations of up to $10,000$ individuals, asexuals rapidly invade fully sexual populations (with an overall fixation probability $u \approx 0.8$), if there is a twofold cost of sex. Specifically, the mean fixation probability u in a population of $10,000$ individuals is 0.77 if mutations are deleterious only, and 0.71 if mutations are advantageous and deleterious. With geographic structure, however, a sexual population can resist invasion, and furthermore the critical average F_{st} values required to favour sex generally decreases as N increases (Figure 5.2). The explanation for this effect is presumably that in larger populations, asexuals take longer to fix and so accumulate more deleterious mutants. In addition, loci are more likely to be polymorphic, increasing the benefits conferred by sex by creating fitter genotypes (MARTIN *et al.* 2006). I observe that higher average F_{st} values were required to maintain sex with two-dimensional spatial structure compared to the one-dimensional model. This reflects the fact that asexuals can more readily establish when they are able to spread in multiple directions.

I observed the lowest average F_{st} values needed to maintain sex if there are both advantageous and deleterious mutations present (Figure 5.2(b)), consistent with results showing that this scenario confers the greatest advantage to recombination (Chapter 3). As a result, critical F_{st} values appear to continually decrease if there is a mixture of advantageous and deleterious mutations, for the population sizes investigated here. By separating advantageous alleles from poor genetic backgrounds, the fixation probability of these alleles is increased, raising the fitness of associated sexuals (FISHER 1930; PECK 1994). Despite this advantage, the F_{st} values required to maintain sex are found to be very high. The lowest F_{st} value of 0.298, for demes spread out over a

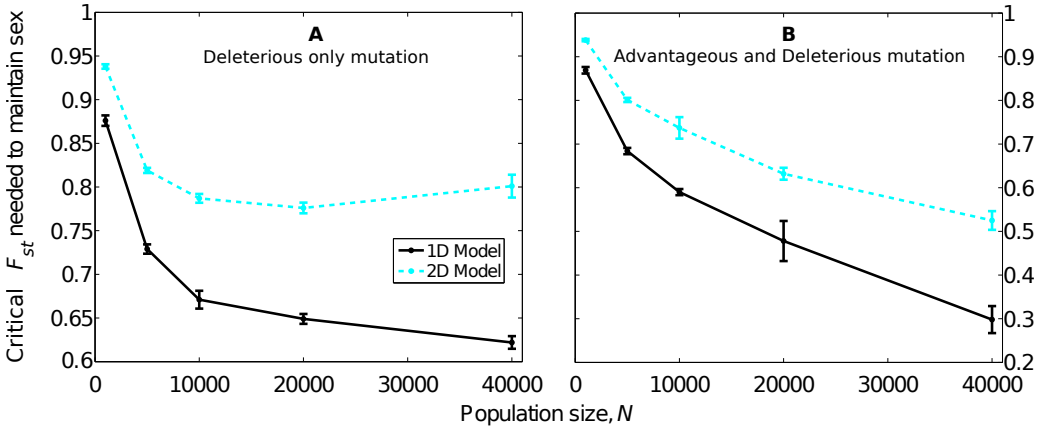


Figure 5.2: The critical level of population subdivision, measured using average pairwise F_{st}, needed to protect a sexual population from asexual invasion, plotted against overall population size, N, with a twofold cost of sex. At the critical F_{st}, asexuals are no more likely to invade than a neutral mutation. Mutations are solely deleterious (a), or advantageous and deleterious (b). Each population is equally spread out over 25 demes, either arranged in a one-dimensional structure (black), or over a two-dimensional torus (light blue). The mutational parameters are: U = 1.0, $s_a = s_d = 0.01$. Confidence limits are based on 1000 bootstraps. If these are not visible then they lie within the plotted points.

one-dimensional formation with advantageous and deleterious mutation, is of the same order to values obtained from plant populations, and eukaryotes spread out over continents (see review of F_{st} estimates in Appendix 5.A). However, surveys of geographically close populations show F_{st} values that are substantially lower than the smallest critical F_{st} obtained.

It can be argued that asexuals have less than a twofold advantage because, for example, sexuals can offset some of their costs due to parental effects (AGRAWAL 2001; SILLER 2001). I therefore next investigated the F_{st} values needed to maintain sex if sexuals only suffered a 1.75 cost. As expected, populations need a greatly reduced level of subdivision to maintain sex, as measured using both average and maximum pairwise F_{st} values (Figure 5.9 in the supplementary figures). For example, if there is a twofold cost, sex is maintained in a one-dimensional population of $N = 20,000$ individuals subject to both advantageous and deleterious mutation if average $F_{st} \approx 0.475$ (Figure 5.2(b)). However, if sexuals suffer a 1.75 cost, then for the same sized population, sex can be maintained if average $F_{st} \approx 0.3$ (Figure 5.9(b)). Note that this is still a high F_{st} value compared to those obtained from natural surveys of population structure.

So far, I have considered models in which individuals are only able to migrate to a neighbouring deme. In natural populations, however, it is expected that individuals migrate across a wide range of distances, and a small proportion of migrants may be able to travel over long distances (WRIGHT 1931; SHIGESADA and KAWASAKI 1997). For example, KARLIN *et al.* (2012) recently found evidence that a single asexual founder of the peat moss *Sphagnum palustre* located in Hawaii has dispersed to long-distance habitats. Therefore, to investigate the effects of long-distance migration, I estimated the critical F_{st} values needed to maintain sex if 5% of migrants could travel to any deme, as opposed to just neighbouring demes. Figure 5.3 shows that whilst levels of F_{st} required to maintain sex decrease with higher population size, as seen in previous results, the levels of subdivision needed to maintain sex are similar for one-dimensional and

two-dimensional populations. Comparison of Figure 5.3 with Figure 5.2 shows that the critical F_{st} needed to maintain sex with long-distance migration is similar to that observed in two-dimensional populations, where individuals only migrate to neighbouring demes.

I also observe that for larger populations, a slightly higher level of F_{st} is needed to maintain sex in one-dimensional populations compared to two-dimensional populations (Figure 5.3(a) and (b)). This is in contrast to models with no long-distance migration, where two-dimensional populations exhibit higher critical F_{st} values (Figure 5.2). I assume that this new behaviour arises because, with long-distance migration in one-dimensional populations, invading asexuals are able to reach distant demes in a much quicker time than if they were only able to migrate to neighbours. Asexuals therefore experience a greater increase in their fixation probability in one-dimensional populations with long-distance migration, compared to two-dimensional populations, as demes are already highly connected in the latter case. To compensate for this, the level of subdivision in one-dimensional populations needs to be greatly increased to maintain sex, so critical F_{st} values exceed those reported for two-dimensional populations.

To investigate whether this pattern holds for lower and higher fractions of long-distance migration, I determined the critical F_{st} for a population of size $N = 10,000$, allowing the proportion of long-distance migrations to vary between 0% (so individuals could only migrate between adjacent demes, as used in previous simulations) and 10%. Figure 5.10 in the supplementary figures demonstrates that irrespective of the type of mutation present, only a small amount of long-distance migration is needed, so that the critical F_{st} values in one-dimensional populations increase and become similar for those reported for two-dimensional models.

Next, I investigated how the critical F_{st} values change if the number of demes varies in simulations of a one-dimensional structured population consisting of $N = 5,000$ individuals overall. Figure 5.4 shows that as the number of demes increases, the critical

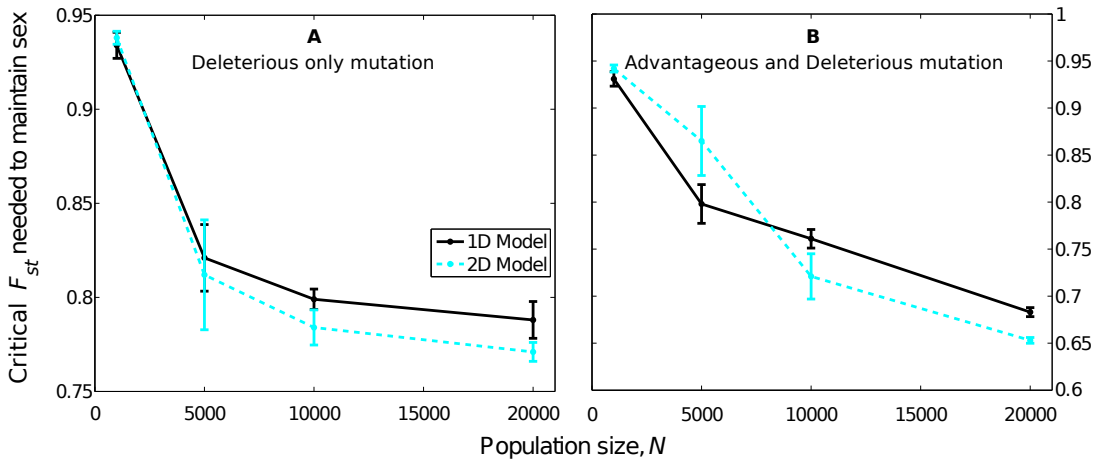


Figure 5.3: The critical level of population subdivision, measured using average pairwise F_{st}, needed to protect a sexual population from asexual invasion plotted against overall population size, N. Here, 5% of migrants travel over long-distances. Mutations are solely deleterious (a), or advantageous and deleterious (b). Each population is equally spread out over 25 demes, either arranged in a one-dimensional structure (black), or over a two-dimensional torus (light blue). The mutational parameters are: U = 1.0, $s_a = s_d = 0.01$. Confidence limits are based on 1000 bootstraps. If these are not visible then they lie within the plotted points.

average F_{st} value needed to maintain sex decreases, implying that sex is more likely to be maintained in species spread over large regions. This result emerges due to the interplay between two major mechanisms. First, since there are more demes, then the invading asexual will take longer to spread through the entire population, as it has to wait longer before being transferred to a new area. Consequently, it is more likely to accumulate deleterious mutation and go extinct (SALATHÉ *et al.* 2006), so lower levels of F_{st} are needed to maintain sex. The second mechanism is that with more demes for a fixed N, each deme consists of a smaller population, which can accelerate the speed at which Muller's Ratchet operates (GESSLER 1995; HIGGINS and LYNCH 2001; GORDO and CAMPOS 2008), further reducing the fitness of asexuals.

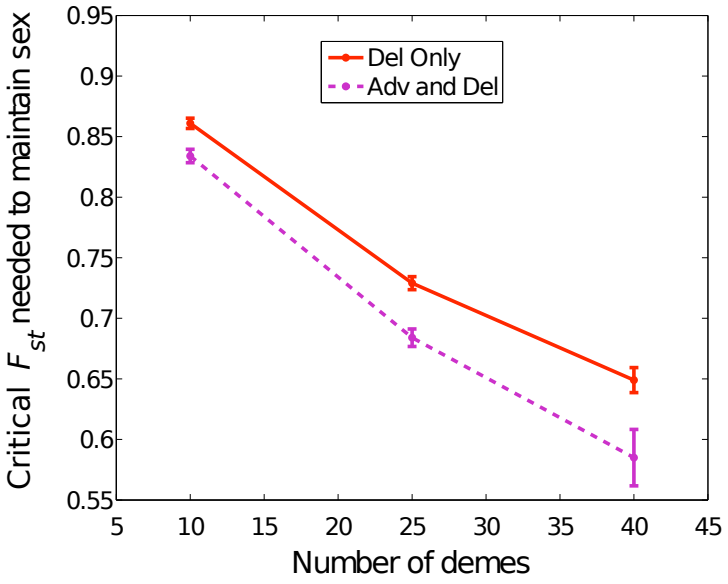


Figure 5.4: The critical level of population subdivision, measured using the average pairwise F_{st}, needed to protect a sexual population from asexual invasion, plotted against the number of demes simulated. Mutations are solely deleterious (red), or advantageous and deleterious (purple). Demes are arranged in a one-dimensional structure. The mutational parameters are U = 1.0, $s_a = s_d = 0.01$. Confidence limits are based on 1000 bootstraps.

5.3.2 The effect of varying the mutation rate on asexual fixation probability

SALATHÉ *et al.* (2006) showed that sex is increasingly favoured as the genomic deleterious mutation rate increases, given that deleterious mutations have a fixed selection coefficient s_d. My results suggest that sex is most strongly selected in the presence of both advantageous and deleterious mutations. To determine to what extent advantageous and deleterious mutations lead to the maintenance of sex, I evaluate the fixation probability of an asexual mutant in the presence of just advantageous mutations or just deleterious mutations. Figure 5.5 shows that asexuals are less likely to establish in populations subject to higher rates of deleterious mutation. If only deleterious mutations are present, the fixation probability of an asexual decreases in an approximately exponential fashion as the mutation rate increases, reflecting an increased likelihood of asexual mutational meltdown (SALATHÉ *et al.* 2006).

The additional advantage to sex conferred by advantageous mutations depends on the advantageous mutation rate. This is still generally unknown for many higher eukaryotes. SHAW *et al.* (2002) estimated that 50% of mutations in *Arabidopsis thaliana* are beneficial (but see KEIGHTLEY and LYNCH (2003)). JOSEPH and HALL (2004) estimated that around 5.75% of mutations in diploid lab-adapted strains of *Saccharomyces cerevisiae* are beneficial. A follow-up study by HALL and JOSEPH (2010) further found that these proportions are higher for mutations that affect growth rate and sporulation efficiency (12.5% and 20%, respectively), but it was inferred that no beneficial mutations that affect spore viability were present, nor were any found in haploid strains. In *Drosophila melanogaster*, SCHNEIDER *et al.* (2011) estimated that the proportion of advantageous amino-acid-changing mutations lies between 0.5% and 3.5%. The value of 2.2% used in these simulations thus lies within this range of estimates.

Figure 5.5 shows the dependancy of the fixation probability of an asexual mutant on the advantageous mutation rate. If only advantageous mutations are present, asexu-

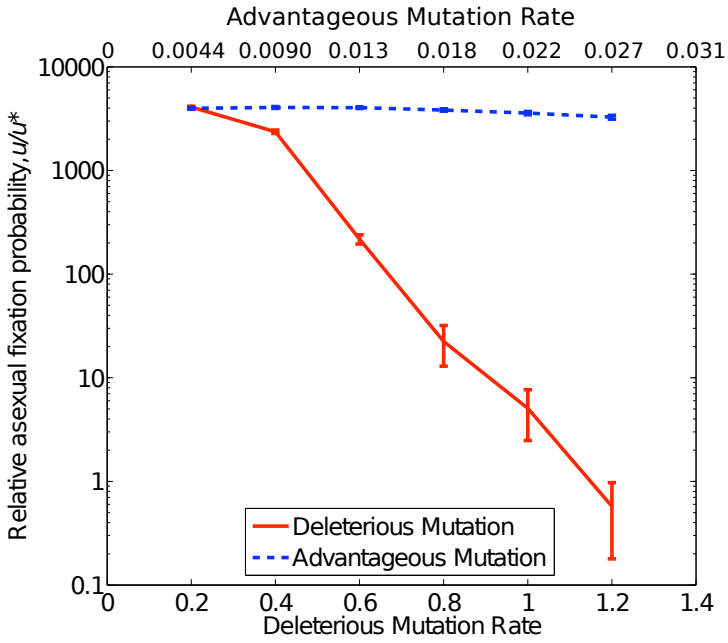


Figure 5.5: Relative fixation probability u/u^* of an asexual invader, plotted as a function of the mutation rate U (see methods section for how U is defined for different mutation schemes). Mutations are deleterious only (red) or advantageous only (blue). N = 5000, $s_d = s_a = 0.01$, $m = 0.003$ (yielding an average F_{st} value of 0.687 with only deleterious mutations for U = 1). 25 demes are spread out over one dimension.

als always fixed at high rates. I conclude that advantageous mutations, acting on their own, are ineffective at protecting structured sexual populations from asexual invasion, at least for the low frequencies at which they arise in these simulations. Nevertheless, combined with deleterious mutations, the presence of advantageous alleles greatly reduces the amount of population structure needed to resist asexual invasion in larger populations (Figure 5.2). This is highlighted in Figure 5.11 in the supplementary figures, which plots the asexual fixation probability with different advantageous mutation rates, if advantageous and deleterious mutations are both present. As the proportion of advantageous mutants decreases, then the consequent increases in asexual fixation probability are much smaller compared to values observed when advantageous mutations arise on their own (compare Figure 5.11 with Figure 5.5).

5.3.3 Effects of varying selection strength of deleterious and advantageous mutation on the fixation probability of an asexual

The previous simulations explored the invasion of asexual variants assuming a single value for the fitness effects of mutations ($s_a = s_d = 0.01$). I next explored how the value of selection influences these results. Although the average selection against deleterious alleles is on this order, the distribution of fitness effects is very broad, with a leptokurtic distribution (EYRE-WALKER and KEIGHTLEY 2007). Advantageous mutations also have varying effects; SATTATH *et al.* (2011), for example, estimated that in *D. simulans* a small proportion of substitutions are strongly selected for (mean $s_a \approx 0.005$), whereas the rest have a smaller effect (mean $s_a \approx 4 \times 10^{-5}$). It has been previously observed that the fixation probability of a recombination modifier depends on s_a (Chapter 3), and s_d also affects the maintenance of sex in a structured population (SALATHÉ *et al.* 2006). Therefore, I next examine how different strengths of mutation, both advantageous and deleterious, affect the maintenance of sex.

If there are deleterious mutations only, or with a mixture of advantageous and de-

leterious mutations, these results suggest that asexual invasion is least likely to occur for intermediate values of the selection coefficient against deleterious mutations (Figure 5.6). This is a consequence of Muller's ratchet being more likely to cause the extinction of asexual lineages for intermediate selection strengths, because weakly deleterious alleles have little effect on asexual fitness while strongly deleterious alleles are unlikely to establish (GABRIEL *et al.* 1993). If the population size is altered, Figure 5.12 in the supplementary figures shows that, whilst there is no noticeable difference to the results for $N = 10,000$ compared to $N = 5000$ (Figure 5.12(b)), there exists a wider range of s_d values for which the asexual does not fix for $N = 1000$ (Figure 5.12(a)) due to stronger effects of Muller's ratchet caused by decreased population size.

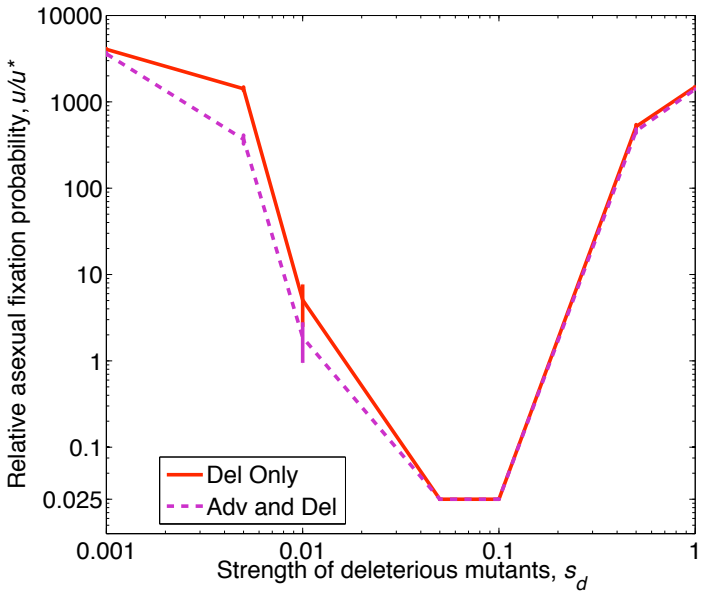


Figure 5.6: Relative fixation probability u/u^* of an asexual invader, plotted as a function of the strength of selection acting against a deleterious mutant s_d. Mutations are either deleterious only (red) or advantageous and deleterious (purple) with $s_a = 0.01$. Other parameters are $N = 5000$, $U = 1.0$ and $m = 0.003$ (yielding an average F_{st} value of 0.687 with only deleterious mutations, for $s_a = s_d = 0.01$). 25 demes are spread out over one dimension. Fixation probabilities of zero were replaced with 0.025, representing a single fixation event over all replicates.

Finally, increasing the selective advantage of advantageous alleles increases the ability of sexual populations to be maintained (Figure 5.13 in supplementary figures).

5.4 Discussion

In recent years, breaking apart Hill-Robertson interference between linked loci generated by selection in finite populations has become a strong candidate for the driving force favouring the evolution of recombination (BARTON 2010). However, less is known about whether this advantage of recombination can overcome strong costs of sex in subdivided populations, with low F_{st} levels that one expects to find in nature. My results in this chapter are consistent with previous studies showing that obligate sexuals suffering a twofold cost are able to resist invasion by asexuals in structured populations (PECK *et al.* 1999; SALATHÉ *et al.* 2006). Furthermore, the amount of structure needed to maintain sex, as measured using the average pairwise F_{st} value between demes, decreases with population size. However, the reported average F_{st} values needed to maintain sex are very high compared to those observed in nature. In the largest populations simulated ($N = 40,000$), the lowest critical average F_{st} observed where sexuals suffer a twofold cost was $\sim$0.298, for demes spread out over one dimension and individuals subject both to advantageous and deleterious mutation. Such a level of population subdivision has been observed in geographically close plant populations, but is greatly higher than values observed in natural populations of fish and mammals (see Appendix 5.A). These results hold for a realistic eukaryotic genome-wide deleterious mutation rate, $U = 1.0$ (DENVER *et al.* 2004; HAAG-LIAUTARD *et al.* 2007; EÖRY *et al.* 2010).

Whilst sexuals are maintained predominantly because asexuals undergo deleterious mutational meltdown, which can lead to their extinction in a metapopulation (HIGGINS and LYNCH 2001), I also observe that sex is most strongly favoured in the presence of both advantageous and deleterious mutations. In the absence of deleterious muta-

tion, critical levels of F_{st} level off at values that are far higher than those seen in nature (Figure 5.2(a)). This finding suggests that, not only do advantageous mutations aid the maintenance of sex, but their presence is necessary for sex to be maintained if levels of population subdivision are low. With advantageous and deleterious mutation the critical average F_{st} falls steeply with increases population size N (Figure 5.2(b)). A likely cause of this phenomena could be the amount of variation generated by background selection and selective sweeps respectively, in the limit of free recombination (this mechanism is discussed in more detail in OHTA and KIMURA (1975)). If each biallelic locus acted independently, the amount of fitness variance it would contribute would be equal to $p(1-p)s^2$ (BULMER 1980). If mutation were deleterious, the frequency of the deleterious allele p would be approximate equal to μ/s (where μ is the per-locus mutation rate), so the variance that the specific locus would contribute equals μs. Over all deleterious loci in the genome, the total variance contributed is equal to Us, which is independent of the population size N. However, advantageous mutations will continue to contribute to genetic variance as they segregate in the population; the number of new advantageous mutations introduced per generation proportional to $N_e U_a$, for U_a the advantageous mutation rate and N_e the effective population size. It is therefore feasible that critical F_{st} values would reduce to realistically low levels if the population size was much larger than those simulated for this study, in the presence of advantageous and deleterious mutations, since there exists more opportunity for genetic variance to be contributed through higher numbers of adaptive mutations. This mechanism favours selection for recombination in large, finite populations (BARTON and OTTO 2005; see also Chapter 3), which can also potentially maintain costly sex in larger populations. I also find that levels of F_{st} needed to maintain sex decreases if asexuals have a less than twofold advantage (Figure 5.9 in supplementary figures), or if populations are spread out over more demes (Figure 5.4).

If a small proportion of long-distance migration is introduced, then it appears that

the critical level of F_{st} needed to maintain sex increases in one-dimensional populations, and is similar to the level of F_{st} needed in two-dimensional populations with no long-distance migration (Figures 5.3, 5.10). This increase in F_{st} in one-dimensional populations presumably reflects how a much lower overall rate of migration is needed to maintain sex, to counteract the increased speed at which asexuals spread to distant demes with long-distance migration.

Possibilities for extending this chapter. Because these simulations are computationally intensive, I have been limited in exploring the parameter space in this complex multidimensional problem. In particular, I have assumed that mutants have equivalent fitness effects, whereas in reality mutants have a distribution of selective effects (EYRE-WALKER and KEIGHTLEY 2007). However, by exploring the impact of different fitness effects (Figures 5.6, 5.12, 5.13) I have obtained some information on the parameters that maximise the likelihood of the maintenance of sex. In particular, deleterious mutations of intermediate effect offer the greatest protection to sex, because these cause the fastest degradation in asexual fitness due to Muller's ratchet. Similarly, haploid populations were simulated, but it is known that selection on sex and recombination acts somewhat differently in diploid individuals. Specifically, if deleterious mutants are strongly recessive, then this can select against increased levels of sex and recombination in diploid populations subject to just deleterious mutation (ROZE 2009; ROZE and MICHOD 2010). Thoroughly investigating how diploidy affects the maintenance of sex in structured populations would be worthy of future study.

I also invoked specific assumptions about the ecology of the simulated populations, which only cover a small part of all biologically realistic scenarios. For example, demes could be subject to 'hard selection' (WALLACE 1975), where the contribution to other subpopulations of offspring from a deme depends on the mean fitness of the individuals within it. Another scenario not considered is where there exists a continuous emergence of asexuals over time, as observed in the system of *Potamopyrgus antipodarume*

snails (DYBDAHL and LIVELY 1995; JOKELA *et al.* 2003; NEIMAN *et al.* 2005; KING *et al.* 2011a), which can lead to parallel fixation of different asexual lineages (RALPH and COOP 2010). This effect might become more apparent in large populations, where there exists a higher probability that multiple asexuals will arise. In higher eukaryotes it is rare for sexuals to mutate to an asexual, so my model can inform on how population subdivision maintains sex in such species. However, investigating whether realistic levels of subdivision can maintain sex in the face of multiple asexuals emerging over time is beyond the scope of this paper, and should be investigated as part of a future study instead.

Throughout the study, F_{st} is used to measure the extent of population subdivision that maintains costly sex. However, whilst F_{st} can determine how population structure increases the fixation time of beneficial alleles (WHITLOCK 2002, 2003), it might not be the most precise statistic to use in determining how population structure maintains sex. For example, the differences in critical F_{st} between one-dimensional and two-dimensional populations disappear with a small proportion of long-distance migration (Figure 5.3). Future theoretical and empirical work should therefore investigate how alternative measures of population subdivision relates to the asexual fixation probability, and whether these remain consistent with different types of population structure. Alternative statistics might include the 'range size' of the metapopulation (the number of demes an asexual would have to travel through before fixing), and the actual number of generations needed for an asexual to fix.

Conclusions. These models have shown that, whilst costly sex can be maintained in subdivided populations, the level of subdivision needed to resist asexual invasion are very high in the population sizes I was able to simulate, compared to levels observed in field studies (Appendix 5.A). The maintenance of costly sex in subdivided populations requires a large number of linked loci subject to both advantageous and deleterious mutation, and large populations spread out over a large number of demes.

However, modelling such a scenario is currently both computationally and mathematically intractable. Therefore it remains to be tested whether costly sex can be maintained in very large subdivided populations with realistic levels of population subdivision.

5.A Review of F_{st} values obtained in studies of natural populations

This appendix provides a review of F_{st} values that have been recorded in nature, in order to ascertain realistic levels of population subdivision. It will focus on several key reviews from the extensive literature on the subject. However some important caveats should be considered with respect to the data presented here. Most modern studies use microsatellite data to measure F_{st}, which tend to give lower values compared to estimates obtained using allozyme data (see, for example, SHAW *et al.* (1999); FREVILLE *et al.* (2001)). Analogous statistics can also be used to measure population subdivision based on microsatellite data, such as R_{st} (SLATKIN 1995). However, traditional population subdivision statistics, especially G_{st}, can be limited by the total homozygosity present in a system (HEDRICK 2005). This can lead to an arbitrary reduction in G_{st} if the mutation rate is high (WHITLOCK 2011), or if a high number of unique alleles are used in a study (JOST 2008).

5.A.1 Fish populations

RIGINOS *et al.* (2011) recently compiled a comprehensive review of benthic teleost fishes, comprising of data from 205 records covering 148 species, to determine how F_{st} is altered by biogeography, egg type and other factors. The review found that average F_{st} values obtained covered the full range of values from zero to one, albeit with a majority of values less than or equal to 0.2. The authors found that pairwise F_{st} posit-

ively correlated with the distance between populations, which is to be expected, but also found that species who laid benthic eggs had higher average F_{st} values than those that laid pelagic eggs.

5.A.2 Mammal and bird populations

Studies of mammal and bird populations tend to find lower levels of genetic structure than in fishes. A comprehensive review can be found in HELLER and SIEGISMUND (2009); the data used in their meta-analysis suggested that G_{st} values found in many mammal and bird populations lies around 0.01 to 0.05. There are some exceptions; a population of European wild boar had a G_{st} value of 0.14, for example.

Several studies have investigated such structure in polar regions. PAETKAU *et al.* (1999) investigated the structure of polar bear populations in the arctic. With 16 microsatellite markers a range of F_{st} values were found after comparing different colonies, lying between 0.002 - 0.108. The authors defined intermediate values of F_{st} as 0.004 to 0.019. On the other hand, ROEDER *et al.* (2001) examined seven microsatellite loci and found no structure amongst colonies of Adélie penguins in Antartica. This was an unexpected find as it was assumed penguins were split into separate colonies, and as such should exhibit high levels of population subdivision. The authors posited that low F_{st} was due to large inter-colony populations, providing adequate scope for genetic variance to be maintained.

Higher levels of structure can be found in sheep populations that do not migrate over large distances. WORLEY *et al.* (2004) examined sheep populations in the Canadian provinces of British Columbia, Northwest Territories and the American state of Alaska. The mean global F_{st} found was 0.160, with values lying in the range 0 and 0.35. A Markov chain Monte Carlo (MCMC) technique was used to estimate that the overall population lied in eight distinct clusters. In contrast, in populations of Scottish red deer a mean $F_{st} = 0.019$ was found, which is significantly higher than zero (99% confidence

limits are 0.015 to 0.022), with four population clusters estimated (PÉREZ-ESPONA *et al.* 2008).

5.A.3 Plant populations

Plants more often exhibit F_{st} values greater than 0.1, as they are more likely to be physically separated, thus restricting gene flow between populations. For example, FREVILLE *et al.* (2001) investigated the spread of *Centaurea corymbosa* over a two square kilometre region of Massif de la Clape in France. Microsatellite data gave a mean F_{st} of 0.23 with a range lying between 0.09 and 0.34.

DUMINIL *et al.* (2007) reviewed data on population structure in plants. Plant F_{st}, as estimated using the G_{st} statistic (NEI 1973), ranged from 0.01 to 0.5, with most values between 0.05 and 0.25. The paper also ascertained whether there was a correlation between G_{st} values and similar plant traits. Statistically significant correlations were found between G_{st} and mating system, perenniality, breeding system and the level of inbreeding. Another review by AGUINAGALDE *et al.* (2005) also found quite high values of G_{st} in woody plant species, with values ranging between 0 and 0.83, with an average value reported of 0.52. Similarly, HAMRICK (1989) (Table 4.4) found a wide range of G_{st} values, ranging from 0.068 to 0.56, from a variety of plant populations. A dataset of HEDRICK (1983) (Table 7.6), obtained from six tree species reports slightly lower values with F_{st} ranging from 0.013 to 0.22, and a mean lying at around 0.1.

ALBERTO *et al.* (2010) recently analysed 12 microsatellite loci from a population of giant kelp residing along the Santa Barbara coast, and found F_{st} values lying between 0.01 - 0.05, which are strikingly lower values than those found in the aforementioned reviews of predominantly terrestrial plants.

5.A.4 *Drosophila melanogster*

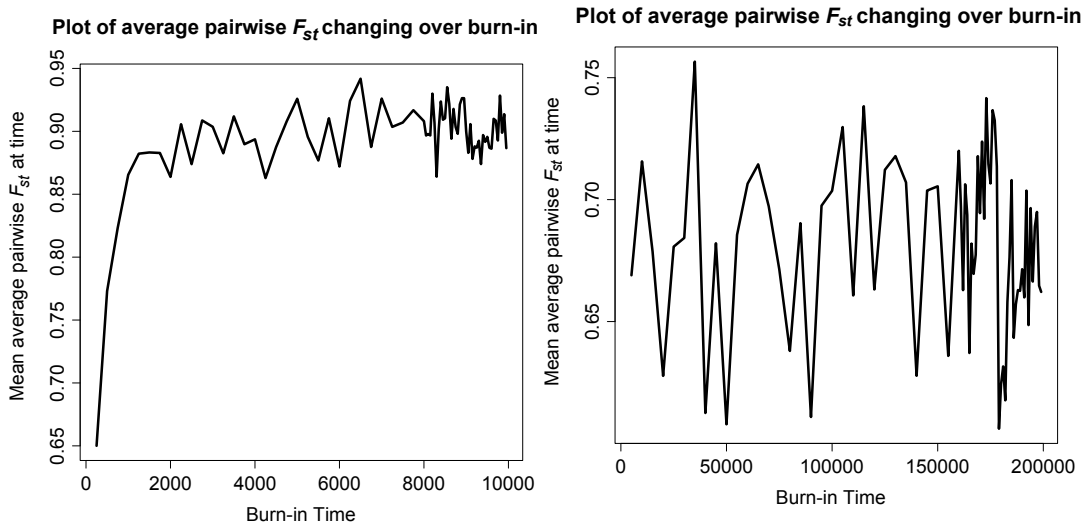
There has also been a solid body of literature investigating population subdivision in *Drosophila melanogster*, mainly concerned with investigating whether the species has a unique African origin. In a general survey, SINGH and RHOMBERG (1987) found that F_{st} values from samples of *Drosophila* taken from around the world were clustered around a mean value of 0.08. However, there was a large tail of high values, extending up to a maximum value of around 0.58. BAUDRY *et al.* (2004) used four microsatellite loci to determine F_{st} in different African and Non-African populations. When comparing populations that both originated within Africa, F_{st} values ranged from 0.02 to 0.1. This increased to 0.15 - 0.45 when comparing Africa and Non-African populations. DIERINGER *et al.* (2005) expanded this analysis by using 17 microsatellite loci (with results confirmed by using an additional 82 X-linked loci). Comparisons of African populations produced F_{st} statistics that were non-significantly different from zero. Non-African populations had a pairwise F_{st} value of 0.03 to 0.1, which increased to 0.13 - 0.3 if an African and Non-African population were compared. Consistent with these results, POOL and AQUADRO (2006) sequenced four X-linked loci and found high $F_{st} > 0.2$ when comparing cosmopolitan and Sub-Saharan populations. Statistically significant F_{st} lying between 0.03 and 0.1 were also obtained when taking pairwise comparisons of African populations.

YUKILEVICH *et al.* (2010) compared North America and African *Drosophila melanogaster* and found $F_{st} \approx 0.1$ when comparing population within America or within Africa. When comparing populations from the two different continents, the authors found mean F_{st} estimates of 0.25 - 0.45 for autosomal loci, but mean values of around 0.5 for X-linked loci, highlighting how different evolutionary pressures can skew estimates of population divergence in different areas of the genome.

5.A.5 Aquatic and marine invertebrates

F_{st} values found in aquatic and marine invertebrates tend to be quite low. MILLER *et al.* (2002) used amplified fragment length polymorphism (AFLP) to investigate population subdivision in four aquatic insect species from the Arizona White Mountains, and found mean F_{st} values lying between 0.01 and 0.06. Similarly in a large review of 25 existing studies of stream invertebrates based on nuclear DNA analysis, HUGHES *et al.* (2008) found that F_{st} values seldom exceeded 0.1 (although a few went as high as 0.2), once potentially cryptic species were removed from the dataset. Recently, in a study of 50 marine invertebrates species from the northeast Pacific area, KELLY and PALUMBI (2010) found that nine non-pelagic species had very high F_{st} with mean values of 0.53. 13 pelagic species had moderate leves of F_{st} (lying between 0.02 - 0.6); the rest had very little population structure.

5.B Supplementary Figures



(a) (b)

Figure 5.7: Example of profile plots for mean average pairwise F_{st} in a population. Data is taken over the burn-in time of $10N$ generations, for $N = 1000$ (a) and $N = 20,000$ (b). F_{st} is sampled more frequently for the last 20% of timepoints. Mutations are deleterious only. There are 25 demes spread out over one-dimension. Migration rate $m = 0.0015$ for both plots.

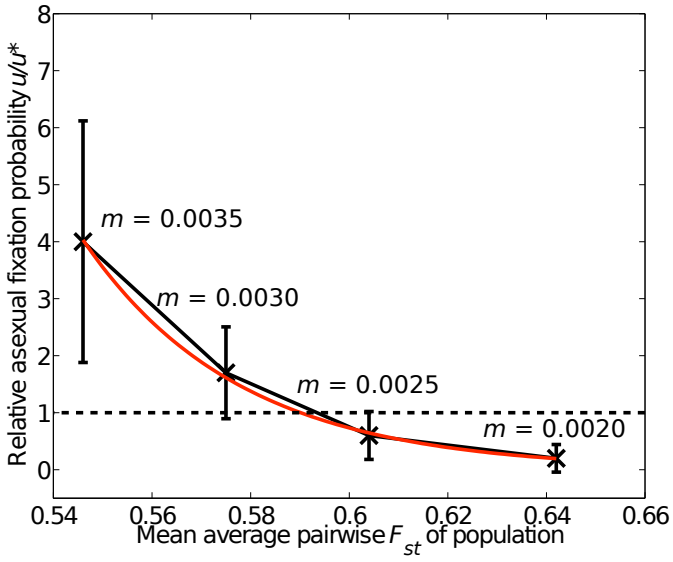


Figure 5.8: A typical profile plot to find the critical average pairwise F_{st} value where sex becomes advantageous. The relative fixation probability of an asexual, u/u^*, is plotted as a function of the mean average pairwise F_{st} recorded over all burn-ins for a particular parameter set (black points). I fitted an exponential curve to the data (red curve), and used this to estimate the F_{st} value where $u/u^* = 1$; that is, where asexuals start becoming selected against. This particular profile is for $N = 10,000$, with advantageous and deleterious mutations present, and demes spread out over one dimension. Bars represent 95% confidence intervals.

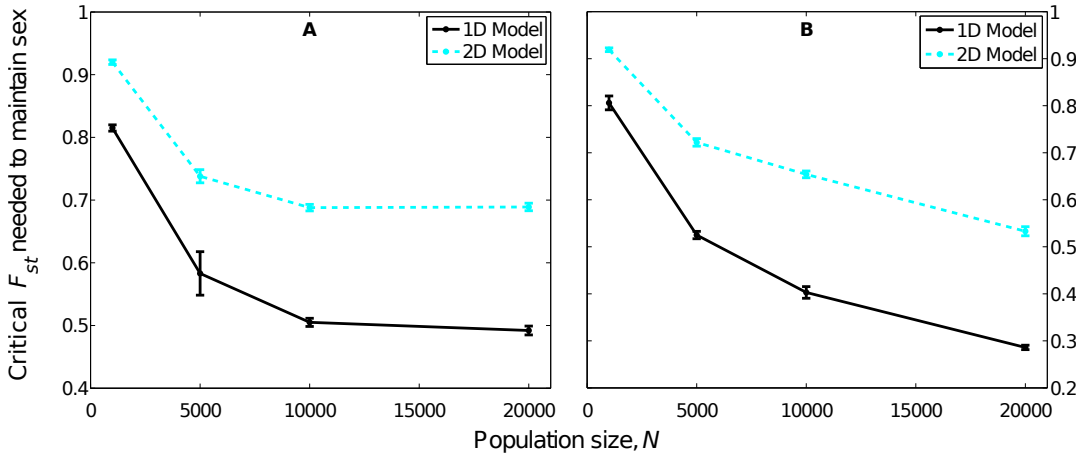


Figure 5.9: The minimum level of average pairwise F_{st} needed to protect a sexual population from asexual invasion plotted against overall population size, N, for a reduced cost of sex $C = 1.75$. Mutations are solely deleterious (a) or advantageous and deleterious (b). Each population is equally spread out over 25 demes, either arranged in a one-dimensional structure (black), or over a two-dimensional torus (light blue). The mutational parameters are: U = 1.0, $s_a = s_d = 0.01$. Confidence limits are based on 1000 bootstraps. If these are not visible, then they lie within the plotted points.

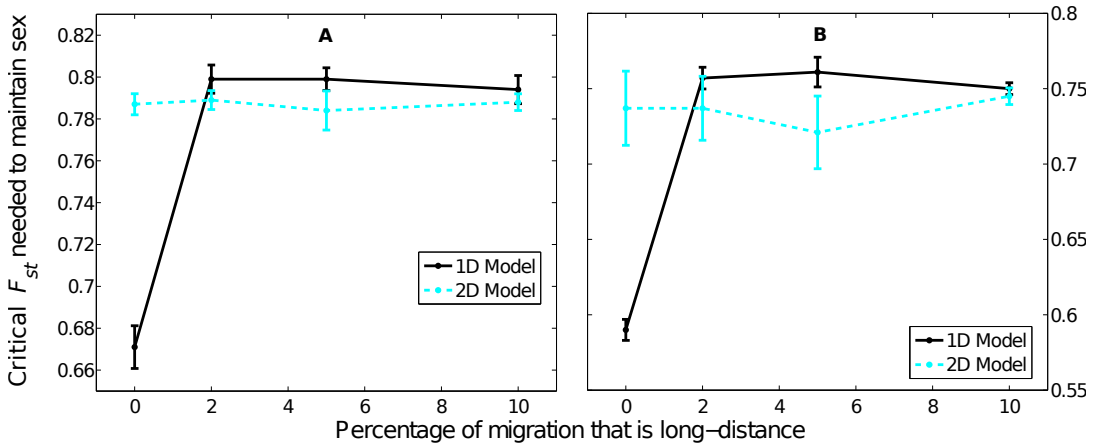


Figure 5.10: The minimum level of population subdivision, measured using average pairwise F_{st}, needed to protect a sexual population from asexual invasion, if there are different proportions of long-distance migration. Mutations are solely deleterious (a), or advantageous and deleterious (b). Each population is equally spread out over 25 demes, either arranged in a one-dimensional structure (black), or over a two-dimensional torus (light blue). $N = 10,000$ and the mutational parameters are U = 1.0, $s_a = s_d = 0.01$. Confidence limits are based on 1000 bootstraps.

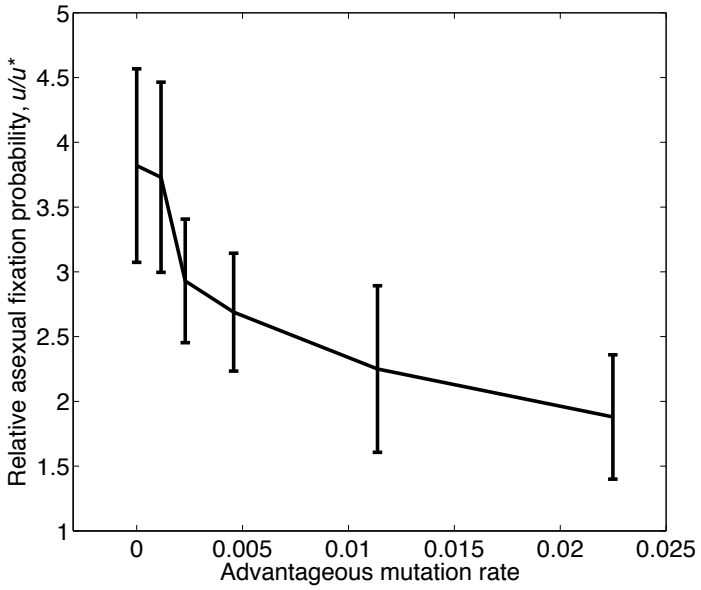
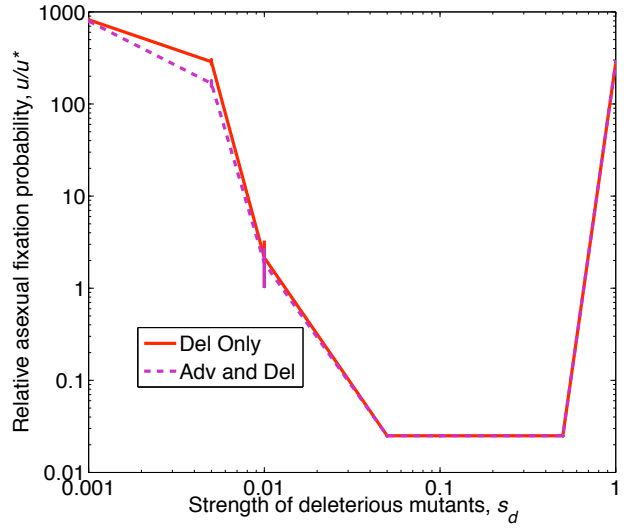
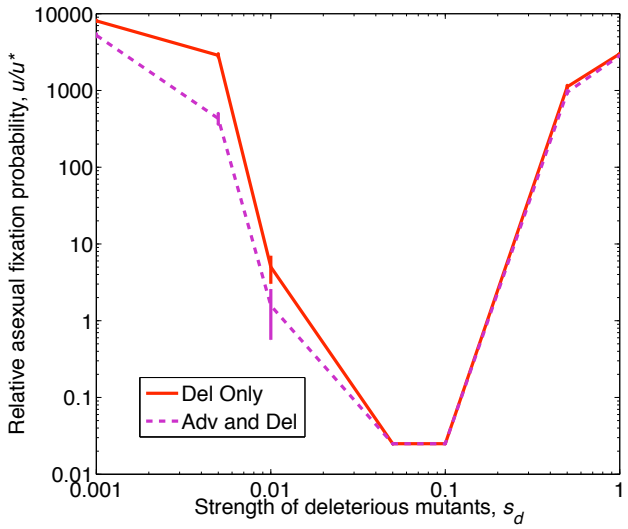


Figure 5.11: Relative fixation probability u/u^* of an asexual invader plotted against the advantageous mutation rate. There was a mixture of advantageous and deleterious mutations. N = 5000, $s_d = s_a = 0.01$, $m = 0.003$ (yielding an average F_{st} value of 0.675 and maximum F_{st} of 0.866 for the baseline advantageous mutation rate, 0.022). 25 demes are spread out over one dimension.



(a)



(b)

Figure 5.12: Relative fixation probability u/u^* of an asexual invader, plotted as a function of the strength of selection acting against a deleterious mutant s_d, for different population sizes. Mutations are either deleterious only (red) or advantageous and deleterious (purple) with $s_a = 0.01$. Other parameters are $U = 1.0$, $m = 0.003$ and (a) $N = 1000$ or (b) $N = 10,000$. These parameters yield F_{st} values of 0.857 and 0.858 for $N = 1000$ and mutations being deleterious only, or advantageous and deleterious respectively for $s_a = s_d = 0.01$; and F_{st} values of 0.613 and 0.575 for $N = 10,000$. Fixation probabilities of zero were replaced with $1/N$.

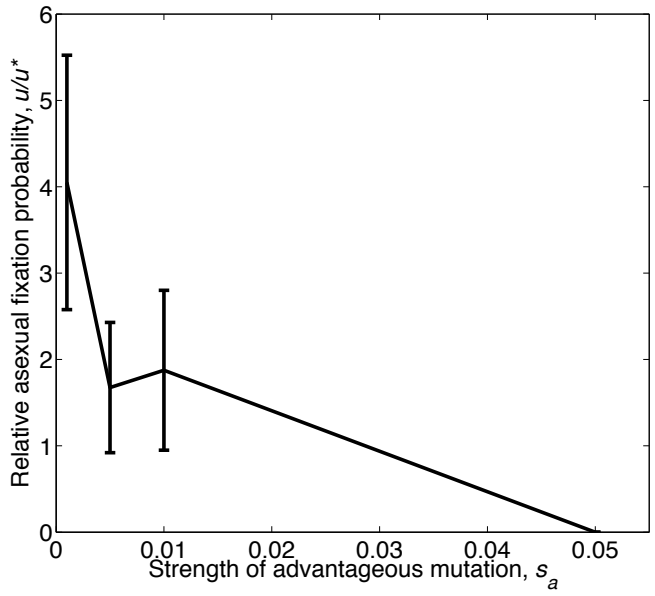


Figure 5.13: Relative fixation probability u/u^* of an asexual invader, plotted as a function of the strength of selection acting on advantageous mutants, s_a, if there was a mixture of advantageous and deleterious mutations. Other parameters are $N = 5000$, $U = 1.0$, $s_d = 0.01$ and $m = 0.003$. Demes are spread out over one dimension.

Chapter 6

A general framework for estimating the fixation time of an advantageous allele in stepping-stone models

Supplementary Material available on attached CD.

Chapter abstract

Determining how population subdivision increases the fixation time of an advantageous allele is an important problem in evolutionary genetics as this influences many processes. Here, I lay out a general framework for calculating the fixation time of a positively selected allele in a subdivided population, as a function of the number of demes present, the migration rate between them, and the manner in which they are connected. Using this framework, it becomes clear that a beneficial allele's fixation time is significantly reduced through migration continuously introducing copies of the allele into a new subpopulation, increasing its frequency within demes. The effect that migration has on allele frequency needs to be explicitly taken into account to produce a realistic estimate of fixation time. This behaviour is most prominent when demes are arranged on a two-dimensional torus, in comparison to populations where demes are arranged in a circle. This is because each subpopulation is connected to several neighbours over a torus, so that there are multiple paths that an allele can take in order to fix. Therefore some demes experience a greater influx and efflux of migrants than others, altering the allele frequency in those specific demes. Analytical results are found to be very accurate when compared to stochastic simulations, and are generally robust if there are a large number of demes, or if the allele is weakly selected for and the migration rate is high.

6.1 Introduction

The interaction between adaptive mutation (HALDANE 1924; FISHER 1930) and population subdivision (WRIGHT 1951) is an area that has been the subject of an extensive body of population genetics research. Much work has focused on how different aspects of population subdivision affect the fixation probability of an advantageous allele (PATWA and WAHL 2008), such as extinction and re-colonisation of demes (BARTON 1993; WHITLOCK 2003; CHERRY 2003b, 2004), the impact of selfing and dominance of mutations (WHITLOCK 2003; ROZE and ROUSSET 2003), frequency-dependent selection (CHERRY 2003a; PANNELL *et al.* 2005), and environmental heterogeneity (LENORMAND 2002; WHITLOCK and GOMULKIEWICZ 2005; VUILLEUMIER *et al.* 2008). There have also been more recent investigations as to how the emergence of multiple advantageous traits interact with each other in spatial populations, and how migration prevents these mutations from interfering with one another (RALPH and COOP 2010; MARTENS and HALLATSCHEK 2011).

One specific area that has attracted interest due to its impact on a wide array of evolutionary phenomena is the fixation time of a favourable allele, as it travels through a series of distinct populations. If the allele has the same selective effect in all subpopulations with additive dominance ($h = 1/2$), and if the deme size is independent of their mean fitness, then the fixation probability of the allele would be the same as in a panmictic population (MARUYAMA 1970). However, the time to fixation will increase if the migration rate $m \ll 1$, which causes an allele to migrate to neighbouring demes in a stepwise fashion.

This slowing effect plays an important role in various evolutionary processes, such as preserving sexuals against asexual invasion (PECK *et al.* 1999; SALATHÉ *et al.* 2006), maintaining underdominant chromosomal inversions (LANDE 1979), altering the dynamics of new species invasion into an existing spatially-extended population, if there is hybridisation with existing species (SHIGESADA and KAWASAKI 1997), and determ-

ining whether migration rates are high enough to prevent neutral divergence between neighbouring regions (MORJAN and RIESEBERG 2004). The slower spread also causes hitchhiking within demes to affect patterns of linked neutral diversity, which can alter measures of population subdivision such as F_{st} (SLATKIN and WIEHE 1998; SANTIAGO and CABALLERO 2005; BIERNE 2010), and skew estimates of the strength of selective sweeps (BARTON 2000; KIM and MARUKI 2011). Substitution rates at selected loci are also reduced, due to the increased time needed to fix adaptive alleles (GORDO and CAMPOS 2006).

FISHER (1937) determined that, if an allele invaded a spatially continuous population, then it would spread with speed $2\sqrt{sm}$, where s is the selective advantage of the allele, and m is the variance in migration distance over an area. This model is accurate if there is a high rate of migration so that the allele travels in a continuous manner ($m \gg s$), and drift effects in the migration rate are negligible ($Nm \gg 1$). It is not applicable in structured populations with low migration rates between adjacent demes, however, as the allele would not spread as a travelling wave. This was demonstrated by SLATKIN (1976), who estimated the mean time taken for a sweep to establish itself in a neighbour, in a two deme system. Using numerical simulations it was found that such a structured population reduces the speed of the spread of the allele by 14-fold, compared to the result predicted using Fisher's travelling-wave solution. SLATKIN (1981) subsequently used Markov-chain methods to estimate an upper limit to fixation time, if migration between demes is weak. KIM and MARUKI (2011) adopted a similar method in their analysis of how population subdivision affects heterozygosity at a linked neutral locus, in a haploid population. They determined that the mean 'delay time' before an allele is established in a new region is given by eqn 5 of KIM and MARUKI (2011):

$$\frac{1}{s} log \left[1 + \frac{s}{m}\right] \tag{6.1}$$

if migration is frequent ($4Nm \gg 1$). A similar result was derived by PIÁLEK and BAR-

TON (1997) when approximating the spread of a travelling wave through a structured population. However, SLATKIN (1976, 1981) and KIM and MARUKI (2011) assumed that the mean time needed for an allele to migrate and establish in a new deme (the 'delay' time) would be the same for every transfer to a new deme that an allele makes, irrespective of the location of the deme or the manner in which it was connected to its neighbours. Therefore, to calculate the overall time needed for an advantageous allele to fix in a population consisting of more than two demes, the mean delay time is multiplied by the number of transfers that the allele makes to a neighbouring deme before it is present in all populations. The analysis in this chapter will show that this assumption is only accurate if migration is very weak ($N_D m \ll 1$ for N_D the population size of the deme, as also determined by SLATKIN (1981)), and subpopulations are arranged in a one-dimensional formation. Otherwise, migration effects will reduce the delay time in subsequent demes. SLATKIN (1976) also assumed that, whilst the rate of spread of an allele would be quicker in a two-dimensional populations, due to the greater number of routes that an allele could take in order to spread, the same lag time would apply to each migration event. It will be shown that the lag times alter between different demes in a two-dimensional population, as some demes experience a greater influx of migrants (and efflux of emigrants) than others.

To try and calculate the fixation time in a more general subdivided population, WHITLOCK (2002, 2003) determined the mean change in allele frequency in a population whose level of subdivision can be measured using Wright's F_{st} statistic (WRIGHT 1951):

$$F_{st} = \frac{V[x]}{\overline{x}(1-\overline{x})} \tag{6.2}$$

where $V[x]$ is the variance in frequency of the selected allele between demes, and $\overline{x}$ is the population mean frequency of the allele. Analytic values were obtained for populations where there is either 'hard' or 'soft' selection. With 'hard' selection, the contri-

bution of each deme to the overall population in the next generation is determined by the mean fitness of the individuals within it, and 'soft' selection arises when each deme contributes individuals independently of the mean fitness of it (WALLACE 1975; WHITLOCK 2002). The terms for the change in mean frequency and variance in frequency were then inserted in the diffusion equations outlined by KIMURA and OHTA (1969) to calculate the fixation time. It was shown that this method provided an accurate estimate when applied to an island model, and a stepping-stone model with demes arranged in a circle.

This chapter aims to extend and complement previous studies by laying out a framework for calculating the fixation time of an advantageous allele in a general structured population, where the allele travels in a stepwise way between demes. The assumption of a stepwise movement of the allele holds if the migration rate is small ($m < s$ for s the selective advantage of the allele). Models are formulated by considering the total number of demes (and the size of each), how they are connected, and the migration rate between each region. An accurate predictor of the fixation time is made by assuming that the allele increases in frequency within each deme deterministically, but the time that such an allele establishes itself in connected neighbours is influenced in a stochastic manner. A similar mix of deterministic and stochastic equations was used by KARASOV *et al.* (2010) to calculate the fixation time of novel mutations arising at the *Ace* locus in a panmictic *Drosophila* population.

Such a model can be applied to investigating natural systems where using F_{st} might not be the most accurate indicator of how subdivided a population is, and thus how it will delay the spread of a selected allele. This may arise if selection acting on loci skews observed estimates of population subdivision (LEWONTIN and KRAKAUER 1973), which arises if $s > m$, as assumed in this analysis (estimates of F_{st} are approximately the same for selected and neutral loci if $s < m$ (WHITLOCK 2002)). F_{st} estimates may also give incomplete information on how the spread of a selected allele is affected by popu-

lation subdivision, if the selective strength of the allele changes over time. The method described in this chapter is flexible enough so that it can be applied to different kinds of stepping-stone model, which is subsequently demonstrated for two types of stepping-stone populations, spread out over one dimension and two dimensions, respectively. An added advantage of this analysis is that it can be used to inform how migration itself can affect the spread of the allele, by introducing copies of it into neighbouring demes after it has established. This can help determine whether migration rates are sufficiently high between demes in order to prevent neighbouring regions from diverging (as reviewed in MORJAN and RIESEBERG (2004)). It also informs on determining when migration is sufficiently high enough in populations consisting of a large number of demes, so that the selected allele moves as a travelling wave. In such cases, Fisher's solution can then be used to measure fixation time instead.

6.2 First model: migration does not affect allele frequency within demes

The first model considers the growth in frequency of a rare allele, which is determined within each deme just by selection acting on it. It is assumed that migration between neighbouring demes can transfer the allele to a new population, but does not affect the allele frequency within each deme. This process continues until the allele fixes in all demes. The mean time taken for an advantageous allele to establish in a neighbouring deme was derived in a similar fashion by SLATKIN (1976) and KIM and MARUKI (2011), and provides a natural starting point for the present analysis. It is assumed that selection acting on the advantageous allele is strong enough that it sweeps through each subpopulation in a deterministic manner, and also that migration is frequent but weak compared to selection ($N_D s \gg 1$ and $m < s$, for N_D the deme population size), so that the sweep travels through each deme in a stepwise formation. In spite of these

assumptions, it will be shown that these models are robust to small $N_D s$ values unless migration is very weak as well; if $m \gg s$ then they will match up to Fisher's (1937) travelling-wave solution.

Consider a finite haploid population of size N, spread equally over D demes, so there are $N_D = N/D$ individuals per deme. After a new generation is created, a proportion m of individuals migrate to a neighbouring deme. At $t = 0$ an individual in a single deme acquires an advantageous mutant, with selection s acting on it (so the fitness of the individual carrying that allele increases from 1 to $1 + s$). It is assumed that the allele is not lost stochastically and proceeds to increase in frequency within that deme. The frequency of the allele at time t is denoted by $p(t)$, which is given by the logistic growth equation (HALDANE 1924):

$$p(t) = \frac{p_0 e^{st}}{1 - p_0(1 - e^{st})} \tag{6.3}$$

Here p_0 is the initial frequency of the allele, which is set to:

$$p_0 = \frac{e^{-\gamma}}{2N_D s} \tag{6.4}$$

where $\gamma \approx 0.577$ is Euler's constant (BARTON 1994). This value is the 'effective' initial frequency, which takes into account the accelerated rise in allele frequency if we only consider cases where the allele is not lost stochastically.

At time t in this first deme, the probability that an advantageous allele migrates to a neighbour is given by $mp(t)$. Once it transfers, the probability of it then establishing itself in the new population is given by $2s$, for $1 \gg s \gg 1/N$ (HALDANE 1927). Thus the overall probability that an allele will migrate and establish itself in a neighbouring deme at that generation is $P(t) = 2smp(t)$. Since an allele only has to establish itself once, then it would have failed to do so in previous generations, each time with probability $1 - P(t')$ (for $t' < t$). Therefore the probability that the first establishment occurs

at time t, denoted by $Q(t)$, is equal to:

$$Q(t) = \left[\prod_{t'=0}^{t-1}(1 - N_D P(t'))\right] N_D P(t) \tag{6.5}$$

Note that $P(t)$ is multiplied by N_D as $N_D m$ is the mean number of migrants between demes every generation. The calculation of eqn 6.5 can be greatly speeded up by approximating the product term; this method was similarly used in simplifying eqn 3 of HARTFIELD and OTTO (2011). If each probability $P(t)$ is small, then the product term can be written as:

$$\prod_{t'=0}^{t-1}(1 - N_D P(t')) \approx exp\left(-\int_0^{t-1} N_D P(t')dt'\right) \tag{6.6}$$

This is a valid approximation since $N_D m$ is not generally found to be large; (MORJAN and RIESEBERG 2004) notes that most estimates from natural populations lie below 10. Therefore, the the compound parameter $N_D P(t) = 2N_D smp(t)$ is small due to the $sp(t)$ term. By evaluating the integral in eqn 6.6, the following is obtained:

$$exp\left(-\int_0^{t-1} N_D P(t')dt'\right) = \left[\frac{2e^{\gamma} N_D s}{2e^{\gamma} N_D s + e^{s(t-1)} - 1}\right]^{2N_D m} \tag{6.7}$$

This derivation is outlined in Supplementary Material 1. From $Q(t)$, the mean time until the allele establishes itself in a neighbouring population can then be calculated. We define this time as $MT1$ ('mean time 1'):

$$MT1 = \sum_{t=1}^{\infty} tQ(t) \tag{6.8}$$

$MT1$ is calculated numerically by computing the sum up to a large upper bound, so that it does not increase further.

In this first model, it is assumed that the rise in frequency of the allele in new demes is determined entirely by selection acting on it, and the effect of migration on its fre-

quency within subsequent demes (through the transfer of alleles between demes) is negligible. Therefore in this model, $MT1$ not only determines the mean time taken for the allele to become established in the neighbouring deme to where the allele first arose, but also other demes thereafter, as assumed by SLATKIN (1976) and KIM and MARUKI (2011). Once the allele establishes itself in the furthest deme, it no longer has to migrate so it only remains to consider the time needed for it to fix within this last deme. Labelling this time as $MT2$, this is given by the time needed for the allele to reach a frequency of $1 - p_0$, at which point it is assumed that the adaptive allele would have almost fixed in the final deme, in a finite population (see also STEPHAN *et al.* (1992)):

$$MT2 = \frac{1}{s} Log \left[\left(1 - \frac{1}{p_0} \right)^2 \right] \tag{6.9}$$

$$= \frac{1}{s} Log \left[(1 - 2N_D s e^{\gamma})^2 \right] \tag{6.10}$$

Note that KIMURA and OHTA (1969) formulated an expression for allele fixation time in a finite panmictic population, using stochastic diffusion equations. However, I use a deterministic equation to calculate $MT2$ so as to retain consistency with the deterministic formulation of $MT1$. Also note that this calculation implicitly assumes that once the furthest deme in the chain has reached fixation then so have all other subpopulations; there are no other demes that are polymorphic at that time. This is a sensible assumption if alleles are strongly selected for, but can be violated for small $N_D s$ values. Despite these caveats, it will be seen that the following models provide an accurate match to simulation data, given these assumptions.

Let there be D' demes between the first deme where the allele first appears and the furthest deme from it. Note that D' is usually not equal to the total number of demes present in a population. For example, if there are D demes arranged in a circular stepping-stone formation, then $D' = D/2$ if D is even, or $D' = [(D-1)/2 + 1]$ if D is

odd. D' signifies the number of demes an advantageous allele has to traverse before it covers the whole population. In this model, after it first appears the advantageous allele will migrate $D' - 1$ times in order to get to the furthest deme, with the mean time taken for each establishing migration to occur equal to $MT1$. Then it has to fix in the furthest deme, which takes $MT2$ generations on average. Thus the mean time to fixation over the whole population is equal to $(D'-1)MT1 + MT2$ generations. Supplementary Material 2 outlines *Mathematica 8.0* code (WOLFRAM RESEARCH 2010) for calculating this value.

The above formulation can be adjusted if the selection coefficient of the allele, or the migration rate differs between demes. $P(t)$, for example, would be altered to become $2s_1mp(t)$, where s_1 is the selection coefficient of the allele in the new deme, and $p(t)$ is calculated using the selective strength of the allele in the original deme. $MT1$ would then have to be recalculated for each separate migration event.

6.3 Second model: migration affects the allele frequency within demes

It is entirely feasible that whilst the advantageous allele is travelling between the first and furthest deme, migration can affect the frequency of the allele in intermediate demes. This situation can arise, for example, through the introduction of more copies of the allele from the previous deme, or if the frequency of the allele is reduced as individuals leave. Since this process can have a significant effect on the fixation time, the first model is adjusted to take such migration effects into account. The basic derivation assumes a fixed s and m, but these can be altered if applying the model to a population with differing values between demes.

In the first deme, migration cannot bring in any new alleles from neighbours, so selection alone determines the frequency of the allele in that deme. Thus the mean time

for it to become established in a second deme is $MT1$, as before. Similarly the time to fixation in the furthest deme is kept as $MT2$. In intermediate demes (demes 2 to $D' - 1$), once the advantageous allele establishes itself, it is assumed that the allele frequency not only changes due to selection, but also due to migration moving copies of the allele between neighbouring demes. In order to account for these extra effects, a system of differential equations needs to be formulated in order to model migration affecting the frequency of the allele within demes. These equations can then be used to calculate the delay time before an allele establishes in a new area, in a similar manner to the first model. As it shall be seen, incorporating these effects into the model causes a significant reduction in fixation time, because, even though migration is weak relative to selection ($m \ll s$), the scaled rate of migration can be significant ($N_D m = O(1)$) and thus can affect the frequency of the allele within different demes.

The simplest way to account for migration effects over a large number of demes is to break the problem down, and consider a closed system of equations in which the allele moves between just two linked demes. These two regions are representative of the deme in which the advantageous allele previously resided, and the deme in which it has just become established. This system therefore assumes that only one other subpopulation 'feeds' advantageous alleles into the current deme; this assumption may be violated if a deme is connected to many neighbours, such as in a two-dimensional torus. The next section demonstrates how migration to and from multiple neighbours can be accounted for.

Define $p_2(t)$ as the frequency of the advantageous allele in the deme where it has just become established. Time is reset so that $t = 0$ is defined as the time when the establishing mutation first appears in the new deme. Furthermore, $q_2(t)$ is defined as the frequency of the allele in the previous deme, from which the advantageous allele is migrating. Under these assumptions, the following set of differential equations is formed:

$$\frac{dp_2(t)}{dt} = sp_2(t)(1 - p_2(t)) - mp_2(t) + mq_2(t) \tag{6.11}$$

$$\frac{dq_2(t)}{dt} = sq_2(t)(1 - q_2(t)) + mp_2(t) - mq_2(t) \tag{6.12}$$

This system considers the allele growing in frequency within the deme due to selection (as denoted by the $sp_2(t)(1 - p_2(t))$ term, along with its equivalent for q_2); migration introducing the allele from the previous deme to the current deme (denoted by the $\mp mp_2(t)$ terms); and migration moving the allele back to the previous deme (denoted by the $\pm mq_2(t)$ terms). Note that in order to keep the system of equations closed (so that p_2, q_2 can reach a maximum frequency of one) we only consider migration occurring between these two demes alone. In reality, migration can also shift copies of the allele back to other demes, or forward to demes where it has yet to establish (such individuals are then lost by stochastic drift). In order to fully account for these migration effects it would be necessary to set up a system of equations for all demes in the chain, which would be unwieldy. However it is possible to produce an accurate model even if these effects are not considered, as they have a minimal effect on allele frequencies. This is because the allele would have fixed in previous demes, so migration from the first deme considered (where the allele frequency is denoted by q_2) to the one that lies previous to it in the chain would not affect the average gene frequency within the first deme. Similarly, only a tiny fraction of individuals would be lost stochastically due to extra migration from the second deme considered (where the allele frequency is denoted by p_2). It will be seen that the adjusted model formed using the above equations still gives an accurate calculation of fixation time.

This system has initial frequency $p_2(0) = p_0$ (as defined by eqn 6.4) and $q_2(0) = p(MT1)$ (the frequency of the allele in the previous deme, at the mean time when it establishes itself in the new population). Such a system can be evaluated numerically

(for example, by using the 'NDSolve' function in *Mathematica*).

Similar calculations as before can be used to find the mean time before the allele establishes itself in a subsequent deme. The probability that an establishing migration event occurs at time t is $P_2(t) = 2smp_2(t)$. As with the previous model, if the first establishing migration occurs at time t, then the allele would have failed to establish in previous generations with probability $(1 - P_2(t))$. So the probability that the first establishing migration takes place at generation t is:

$$Q_2(t) = \left[\prod_{t'=0}^{t-1} (1 - N_D P_2(t'))\right] N_D P_2(t) \tag{6.13}$$

Therefore the mean time for establishment in the next deme is defined as:

$$MT1a = \sum_{t=1}^{\infty} tQ_2(t) \tag{6.14}$$

In this second model, the allele takes $MT1$ generations to leave the first deme and establish itself in the second. It then takes $MT1a$ generations, on average, for the allele to establish itself in subsequent demes, which occurs $D' - 2$ times if s, m do not differ between demes. Finally the allele fixes within the furthest deme in $MT2$ generations. So under this model the mean number of generations needed for the allele to fix would be $MT1+(D'-2)MT1a+MT2$. Supplementary Material 3 gives an example notebook that calculates this time.

Weak migration approximation – In the limit of weak migration relative to selection ($m \ll s$), it is likely that the allele would be fixed in the preceding deme at the time when it establishes in the focal deme. In this case, by setting $q_2 = 1$ in eqns 6.11–6.12, a single differential equation is produced:

$$\frac{dp_2(t)}{dt} = sp_2(t)(1 - p_2(t)) + m(1 - p_2(t)) \tag{6.15}$$

This can easily be solved:

$$p_2(t) = 1 - \frac{(m+s)(1-2N_D s e^{\gamma})}{s(1-2N_D s e^{\gamma} - e^{(m+s)t}(1-2N_D m e^{\gamma}))} \tag{6.16}$$

This form of $p_2(t)$ can be used with eqn 6.13 and 6.14 to obtain a weak-migration approximation for $MT1a$. As for model one, it is possible to approximate the product term in eqn 6.13 since each compound probability $P_2(t)$ is small:

$$exp\left(-\int_0^{t-1} N_D P_2(t')dt'\right) = \left[\frac{2(m+s)e^{\gamma - m(1-t')}N_D}{2N_D s e^{\gamma} + e^{-(m+s)(1-t')}(1+2N_D m e^{\gamma}) - 1}\right]^{2N_D m} \tag{6.17}$$

Together with eqn 6.14, this approximation can be used to produce an analytical formula for the mean fixation time. This derivation is given in Supplementary Material 4.

6.3.1 Correction for multiple demes in a two-dimensional population

The above derivations are good starting models where the spread of an advantageous allele can be described as a series of sequential migrations to connected demes along a linear path. This assumption holds, for example, if demes are arranged in a circular formation, with migration possible between one of its two neighbouring demes (hereafter denoted as the “one-dimensional" case). However, if there exist multiple paths along which the advantageous allele can travel, given the first deme that it has migrated to, then these approximations overestimate the time taken for an advantageous allele to fix. This situation arises if demes are arranged in a grid over a two-dimensional torus, with migration possible from a deme to one of its four neighbours (hereafter denoted as the “two-dimensional" case).

Without loss of generality, assume that the advantageous allele starts in the centre of the grid in a two-dimensional population, and has to migrate to a deme that lies furthest away from where the allele first arose. For a 3×3 array of demes, there are two possible paths that the allele can take to a specific end-point, with no deviation from each (Figure 6.1(a)). However for a 5×5 array of demes, there can be multiple routes that the advantageous allele can travel along to reach the furthest deme, given the first deme that it migrates to (Figure 6.1(b) shows a sample of these routes).

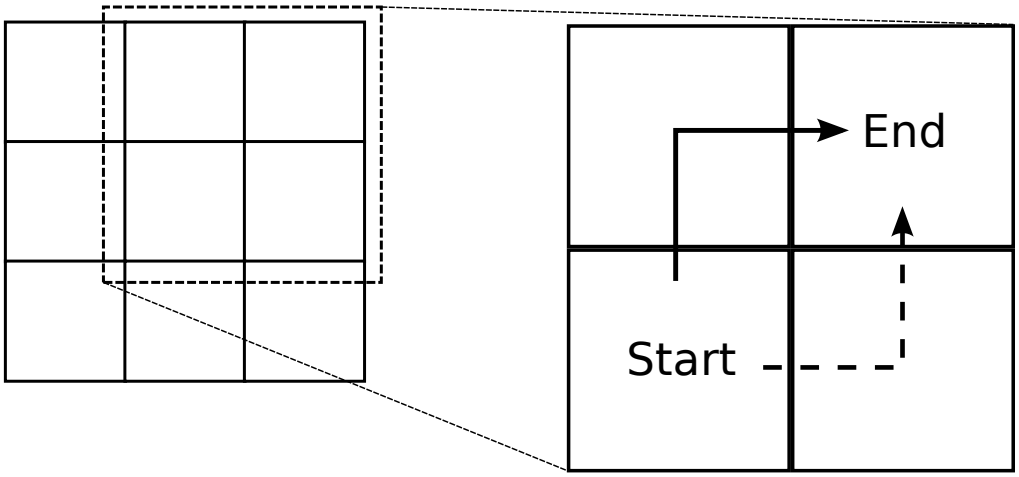
Because of this, the second model needs to be altered to consider these differing migration effects. This derivation is altered in two ways. First, the migration coefficient is scaled to reflect the fact that each deme is connected to more than two neighbours. Second, the multiple routes that an adaptive allele can take to fixation is also taken into account. These points are addressed in turn; Supplementary Material 5 contains example code for implementing these corrections.

Correcting model two to account for multiple neighbours. With populations arranged in a two-dimensional structure, the migration value used in the models had to be set to half that used when applied to one-dimensional populations. This change represent the variance in migration across an individual axis (for migration between adjacent demes), which is half the overall migration rate. Thus eqn 6.11 – 6.12 are rewritten as:

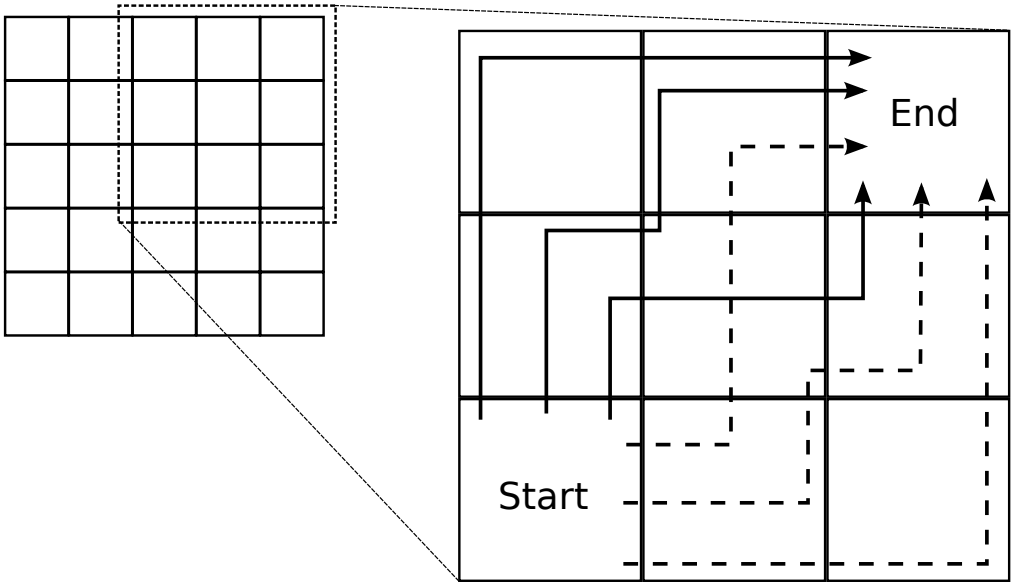
$$\frac{dp_2(t)}{dt} = sp_2(t)(1 - p_2(t)) - \frac{m}{2}p_2(t) + \frac{m}{2}q_2(t) \tag{6.18}$$

$$\frac{dq_2(t)}{dt} = sq_2(t)(1 - q_2(t)) + \frac{m}{2}p_2(t) - \frac{m}{2}q_2(t) \tag{6.19}$$

with the probability that an establishing migration event occurring being equal to $P_2(t) = smp_2(t)$. Similarly, the migration coefficient in Fisher's travelling-wave solution $2\sqrt{sm}$ is scaled by $1/\sqrt{2}$, representing the variance in migration over two dimensions. This is



(a)



(b)

Figure 6.1: Schematic of an advantageous allele travelling from its starting deme, to the furthest-away deme, if subpopulations are spread out over a two-dimensional torus. Schematics are shown for (a) a 3×3 array of demes, or (b) a 5×5 array of demes. Dashed paths indicate the possible paths given that the advantageous allele migrates right initially; solid paths are those where the allele migrates up. For ease, only the paths that do not require the advantageous allele to migrate off the edge of the torus are shown.

discussed further in Appendix 6.A.

Correcting to account for multiple paths to fixation. Because of the multiple paths that an adaptive allele can take when spreading through the entire population, the second model needs to be altered to take these extra routes into account. Each possible path is considered in turn, and for each deme that lies along it the number of possible entrance points and exit points are taken into account in determining how migration affects the frequency of the allele within demes, or the probability of the allele establishing in a neighbour. It will be shown that this adjustment will offer an accurate correction for the population structures considered here, due to the small number of paths considered.

Equations 6.18-6.19 are altered to account for the fact that certain demes experience a greater influx of migrants than others, or that there are multiple demes that the advantageous allele can migrate to, whilst travelling to the furthest point. If there exists a deme on the path in which there exist two possible entrance points for the allele, then we consider migration contributing new copies of the mutation into the focal deme from two preceding demes. As an approximation, the usage of $q_2(t)$ is changed so that in this case it represents the mean frequency of the allele in both these preceding demes, which is equal to the allele frequency in a single deme under the previous model (eqn 6.18-6.19). This is a valid simplification to make if the frequency of the selected allele in both subpopulations is approximately equal to each other at the time when it establishes in the focal deme. This assumption is reasonable, since the alleles spreads in all directions at equal speed, and the selective advantage of the mutant is the same in all demes in this example. Therefore the coefficient of migration used in the equations increases by a factor of two, so for a deme experiencing input of adaptive alleles from two neighbours, eqn 6.11-6.12 are used to model the increase in allele frequency. Similarly, if there are two possible exit points that an advantageous allele can take in order to reach the same end deme, eqn 6.13 is calculated with $P_2(t) = 2smp_2(t)$ instead for that deme, as for

a one-dimensional population. For a two-dimensional grid, these are the only changes that need to be made to the original equations, since no more than two demes can feed advantageous alleles into another deme at any time, nor are there more than two possible neighbours for which an allele can then travel to.

To demonstrate how this correction can be implemented, a 5×5 grid of demes is used as the simplest possible model to which the adjusted equations can be applied. However, it should be noted that, if this correction was to be applied to a system with a larger number of demes, then the following derivation would have to be altered to take into account extra paths that may not be present in this specific example. Nevertheless, it will be shown that a scaled version of this correction is accurate for populations consisting of a large number of demes ($D = 100$) with high migration rates ($N_D m \geq 1$). From Figure 6.1(b) it can be seen that out of the six possible paths, two of them pass through a deme with one possible entrance and two possible exits; one with one entrance and one exit; and a third with two entrances and one exit. Similarly, there are four paths passing through a deme with one entrance and two exits, a second deme with two entrances and two exits, and a third deme with two entrances and one exit. By averaging over all these possible combinations, a corrected form of eqn 6.14 is obtained that accounts for the increased speed at which the advantageous allele spreads at. Let T_a be the mean time taken for the allele to migrate to a neighbour, if present in a deme with one entrance and two exits; T_b the mean time if a deme has one entrance and one exit; T_c the mean time if a deme has two entrances and one exit; and T_d the mean time if a deme has two entrances and two exits. So, for example, T_a is calculated using eqn 6.18 and 6.19 to determine the frequency of the allele at a specific time, then $P_2(t) = 2smp_2(t)$ is used to calculate the probability that it then establishes in a neighbour at time t. By the above reasoning, the mean time taken to migrate in intermediate demes, $MT1a$, is now:

$$MT1a = (2(T_a + T_b + T_c) + 4(T_a + T_d + T_c))/6 \tag{6.20}$$

$$= T_a + T_c + (1/3)T_b + (2/3)T_d \tag{6.21}$$

Note that the above formulation does not take into account paths that wrap around the torus in order to travel to the end deme. However, the results will show that, even without considering these paths the corrected calculation is very accurate, as the ratio of different paths with a certain number of entrance and exit points, as given by eqn 6.21, would remain the same.

6.4 Simulation Methods

In order to test the accuracy of these models, the analytical results were compared to values obtained from stochastic simulations coded in C, which track the spread of an advantageous allele through different types of subdivided population. Simulations start with N haploid individuals divided equally over D demes, with $N_D = N/D$ individuals present per deme. Both one-dimensional and two-dimensional structures are simulated.

A new generation is created according to a Wright-Fisher sampling scheme (FISHER 1930; WRIGHT 1931). Within each deme a parent is randomly selected with probability proportional to its fitness, and then cloned to produce an offspring. This is repeated N_D times so that the whole deme is regenerated, which is then repeated for all demes. Individuals then migrate to neighbouring demes. The number of migrants is chosen from a Poisson distribution with mean $N_D m$. m is the same between each pair of neighbouring demes. For each deme a migrating individual is chosen at random, then moved to a randomly-chosen neighbour. An individual from the neighbour is then moved back to the focal deme, so that N_D is kept constant.

Initially the advantageous allele is introduced into a single, randomly-selected individual in the first deme. The allele increases the fitness of the individual from 1 to $1+s$; s is the same in all demes that the allele resides in. The population then undergoes subsequent selection followed by migration until the mutant is fixed or lost in all demes. If it is fixed, it is noted how many generations it took. This is repeated until the allele fixes 1000 times, so that the mean fixation time with a standard error is produced.

6.5 Model versus simulation results

To test these models, simulations were run with $N_D m$ varying between 0.1 and 5, and $N_D s$ initially varying between 10 and 50. The first results compare the accuracy of the models for a one-dimensional structure with five demes (so the maximal distance $D' = 3$) or eleven demes ($D' = 6$), as well as a two-dimensional structure with a grid of either 3×3 demes ($D' = 3$), or with a 5×5 grid ($D' = 5$). The weak-migration approximation (eqn 6.14 using eqn 6.17) is presented here for one-dimensional populations only. For comparison, simulation results are compared to the fixation time predicted using Fisher's (1937) travelling wave model, $t = D'/(2\sqrt{sm})$. Whilst plots are only shown here for $N_D = 2000$ (except for the cases with a large number of demes), the behaviour outlined below is qualititatively similar for $N_D = 500$ and 1000.

For one-dimensional model results, the second model is very accurate for nearly all $N_D s$ cases for $D' = 3$ (Figure 6.6 in the supplementary figures) and $D' = 6$ (Figure 6.2(a) and (b)), unless migration is weak. For example, if $N_D m = 0.1$ (Figure 6.2(a)) the first model agrees better with simulation data. For all other migration rates investigated (see also Figure 6.7(a) and (b) in the supplementary figures for $N_D m$ = 0.5 and 1 results), model two agrees almost exactly with simulations for all other cases. The exception is $N_D s = 10$, where the first model most agrees with the simulation results for $D' = 3$, and $D' = 6$ with $N_D m = 0.1$ (Figure 6.2(a)). This presumably

arises due to extra stochastic processes caused by weak selection and migration, which significantly affect the fixation time. A full stochastic treatment would be needed in order to account for these. If $N_Dm = 0.1$, the weak-migration approximation generally slightly overestimates the mean fixation time (Figure 6.2(a)). As expected, this approximation greatly overestimates fixation time if $N_Dm = 2$ (Figure 6.2(b)). Fisher's approximation always underestimates the fixation time, especially for $N_Ds = 10$, as the allele does not continuously spread through space, whereas Fisher's model is for an allele spreading through continuous space.

It was also tested whether these models were still accurate if the overall population consists of a large number of demes. Figure 6.2(c) and (d) shows that, with a one-dimensional structure, the second model is accurate if there are 101 demes ($D' = 50$) (see also Figure 6.7(c) and (d) in the supplementary figures for N_Dm = 0.5 and 1 results). As with results for a smaller number of demes, the first model provides the better fit for $N_Dm = 0.1$, although the differences between the two results are very small (Figure 6.2(c)). As N_Ds increases, Fisher's approximation starts overlapping with simulation results, suggesting that the fixation time of the allele can be modelled as a travelling wave in continuous space for these particular paremeters. As with $D' = 6$ the weak-selection approximation slightly overestimates the fixation time if N_Dm = 0.1.

The models are also accurate when applied to a population spread over a two-dimensional torus. For $D' = 3$ (Figure 6.8 in the supplementary figures), simulation data closely match the predictions of model two, again with the exception of $N_Ds = 10$ where both models underestimate simulation results. If $D' = 5$ (Figure 6.3(a) and (b); see also Figure 6.9(a) and (b) in the supplementary figures for $N_Dm = 0.5$ and 1 results), then both models initially overestimate the simulation result, with the exception of $N_Dm = 2$ for $N_Ds = 10$. However, once corrected to account for multiple paths (as outlined in the previous section), model two then agrees accurately with simulations for N_Ds between 20 and 50. Surprisingly, the corrected second model is quite accurate

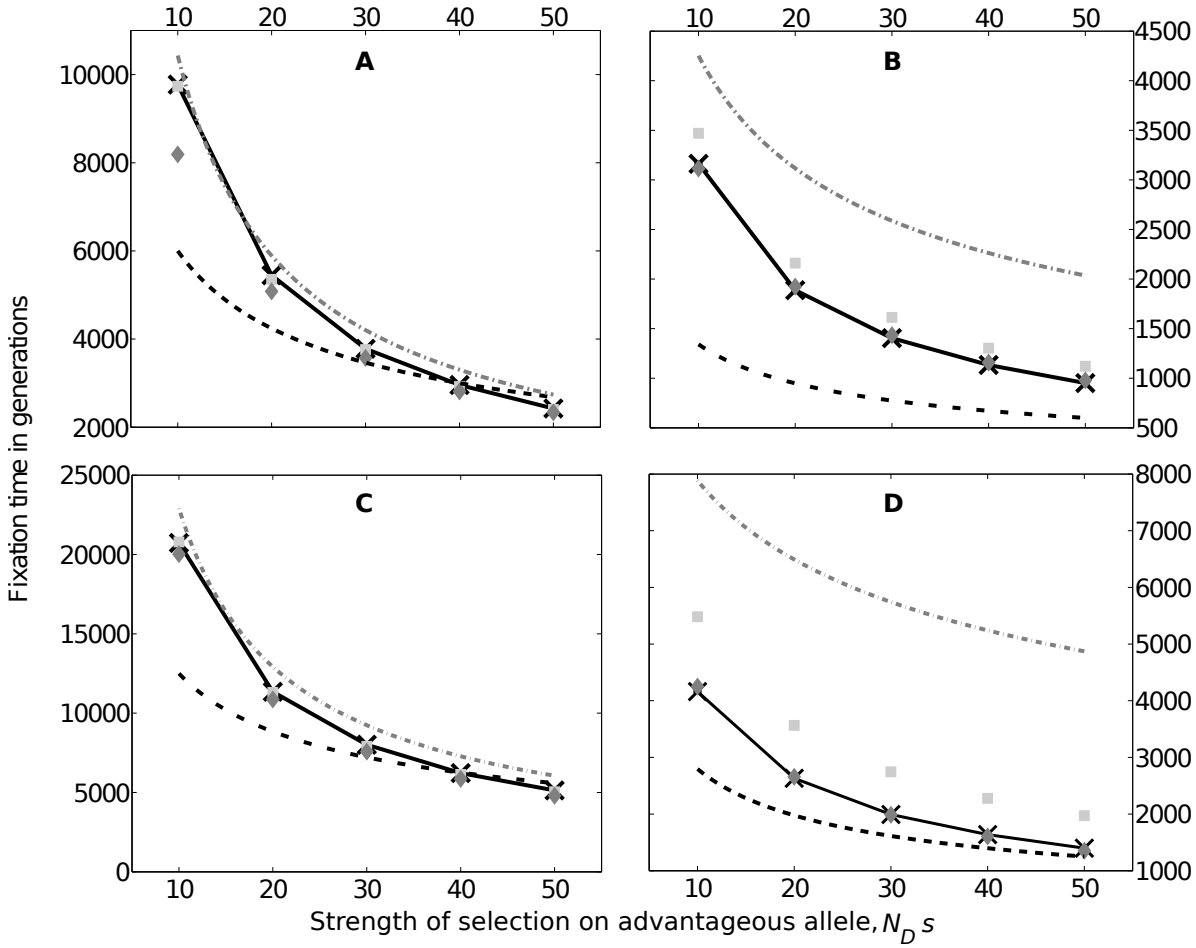


Figure 6.2: Fixation time of an advantageous allele where the population is divided over a one-dimensional structure with 11 demes ($D' = 6$; (a) and (b)), and with 101 demes ($D' = 50$; (c) and (d)). Results are plotted for the first model (light gray squares), second model (dark gray diamonds), weak-migration approximation (dark gray dot-dashed line), simulation results (black crosses joined by a line, standard errors lie within the markers) and Fisher's approximation (black dotted line). $N_D = 2000$ ((a) and (b)) or $N_D = 500$ ((c) and (d)), and $N_D m = 0.1$ ((a) and (c)) or $N_D m = 2$ ((b) and (d)).

for $N_D m = 0.1$ (Figure 6.3(a)), in contrast to one-dimensional results where model one fitted simulation data better for $N_D m = 0.1$ (Figure 6.2(a)).

For a two-dimensional population with 100 demes ($D' = 10$), the corrected form of model two with $MT1a$ (as given by eqn 6.20 and 6.21) had to be scaled by $8/3$, so all the coefficients in eqn 6.21 summed to 8, which is the number of intermediate demes ($D' - 2$). After this change is made, the corrected form of model two is accurate for $N_D m = 2$ (Figure 6.3(d)) and $N_D m = 2$ (Figure 6.9(d)), but significantly overestimates simulation results for smaller migration rates and $N_D s \lesssim 30$ (Figure 6.3(c); see also Figure 6.9(c)). This discrepancy probably arises due the presence of more paths that an advantageous allele can take whilst fixing compared with populations consisting of fewer demes, which are not accounted for in the original derivation.

Next, it was investigated how the accuracy of each model changed with different values of the migration rate, $N_D m$. Figure 6.4 plots the fixation time of an advantageous allele as a function of the migration rate $N_D m$, in populations consisting of a small number of demes ($D = 11$ for one-dimensional models, and $D = 25$ for two-dimensional populations). This was investigated with two different values of $N_D s$ (10 and 50). In one-dimensional models (Figure 6.4(a) and (b)), model two provides a very good match to simulation data for all $N_D m$ values, with the corrected version of model two providing the most accurate match in two-dimensional populations (Figure 6.4(c) and (d)). Exceptions arise if $N_D s = 10$ with a weak rate of migration ($N_D m = 0.1$). In one-dimensional populations, model one provides the best fit to data, whilst all models overestimate the fixation time in two-dimensional populations if $N_D s = 50$. As expected, in one-dimensional populations the weak-migration approximation is only accurate for $N_D m \approx 0.1$ (Figure 6.4(a) and (b)). It is also observed that Fisher's approximation starts to agree with simulation results if the migration rate is low ($N_D m \leq 0.5$), and the allele is strongly selected for ($N_D s = 50$). Otherwise, the analytical models presented in this chapter provide a better match with simulation data. The same behaviour is also

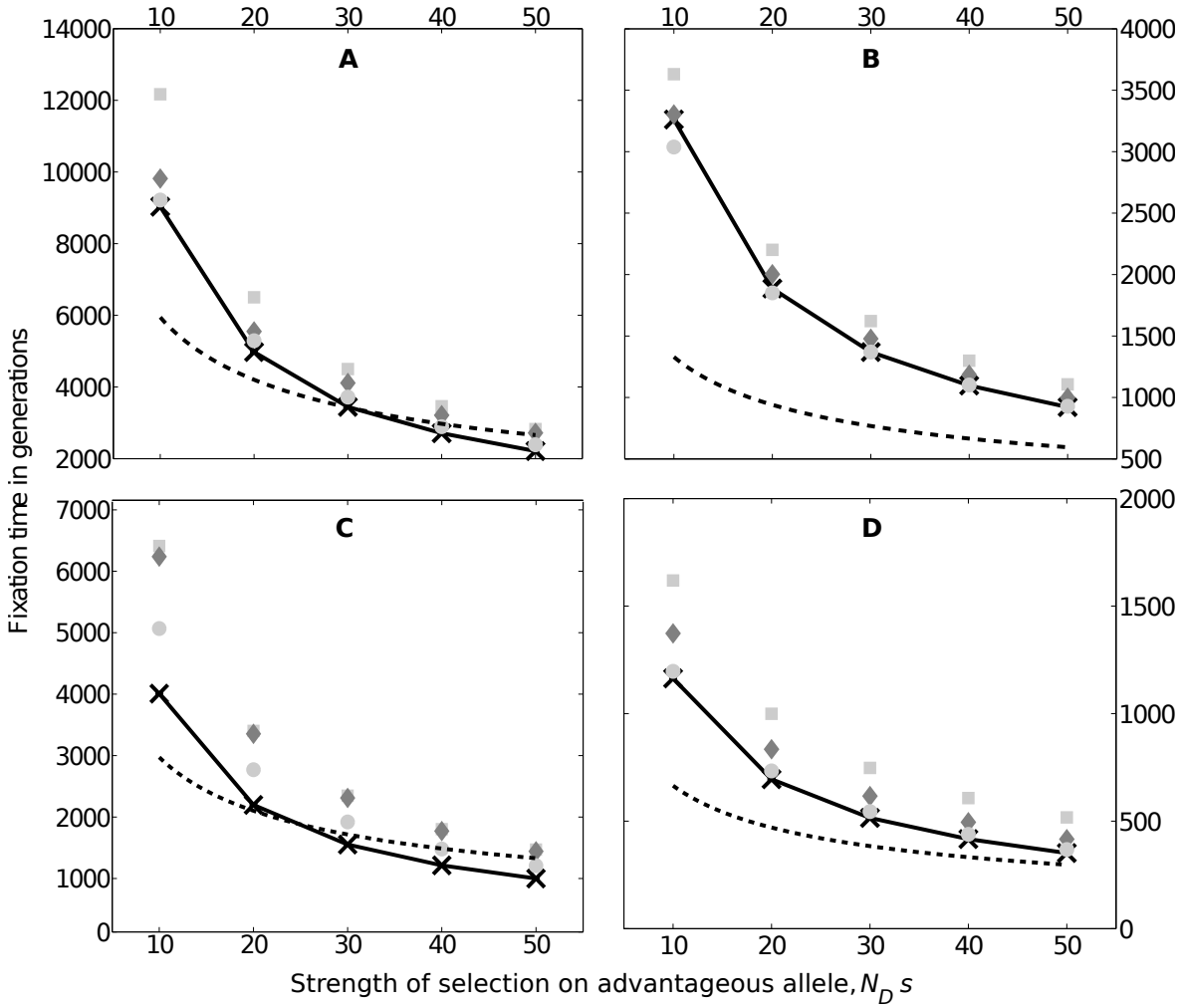


Figure 6.3: Fixation time of an advantageous allele where the population is divided over a two-dimensional structure with 25 demes ($D' = 5$; (a) and (b)), and 100 demes ($D' = 10$; (c) and (d)). As well as plotting the simulation data and two model results, the corrected version of model 2 that accounts for the different ways in which an advantageous allele can reach a target deme is also shown (light gray circles). $N_D = 2000$ ((a) and (b)) or $N_D = 500$ ((c) and (d)), and $N_D m = 0.1$ ((a) and (c)) or $N_D m = 2$ ((b) and (d)).

observed if there are a large number of demes ($D = 101$ for one-dimensional models, and $D = 100$ for two-dimensional populations; Figure 6.10 in the supplementary figures). It was also determined that the second model provides a good match with simulation data for $N_D m = 5$ (see Figure 6.11 in the supplementary figures for plots using different values of $N_D m$), although there is no single accurate model for $N_D s = 10$.

One implicit assumptions of the analysis is that the overall strength of selection is large, so each allele increases in frequency within each deme in a deterministic manner. To test how robust these models are for weak selection, they were compared against simulations with $Ns = 100$, with N representing the overall population size (so that $N_D s = 1$), where there is a large stochastic component determining the frequency of the allele in each deme. As Figure 6.5 shows, both models vastly underestimate the fixation time for $N_D m = 0.1$, presumably because stochastic effects not considered in the analysis strongly affect the allele fixation time. For $N_D m = 0.5$ to 2 the first model agrees well with simulation data. For $N_D m = 5$ both models slightly underestimate the simulation fixation time, and Fisher's travelling wave model agrees best instead. Here, migration is more stronger than selection so that the allele spreads in a continuous manner.

6.6 Discussion

This chapter shows how a mixture of deterministic models representing the increase in allele frequency within demes, combined with a stochastic analysis of the mean time needed for the allele to establish in a new area, can be combined in order to produce an analytical estimate of the fixation time of advantageous allele in a subdivided population consisting of multiple demes. It is shown that the second model outlined here, which takes into account migration altering frequencies of the advantageous allele in intermediate demes, provides a very good estimate of the fixation time for most cases,

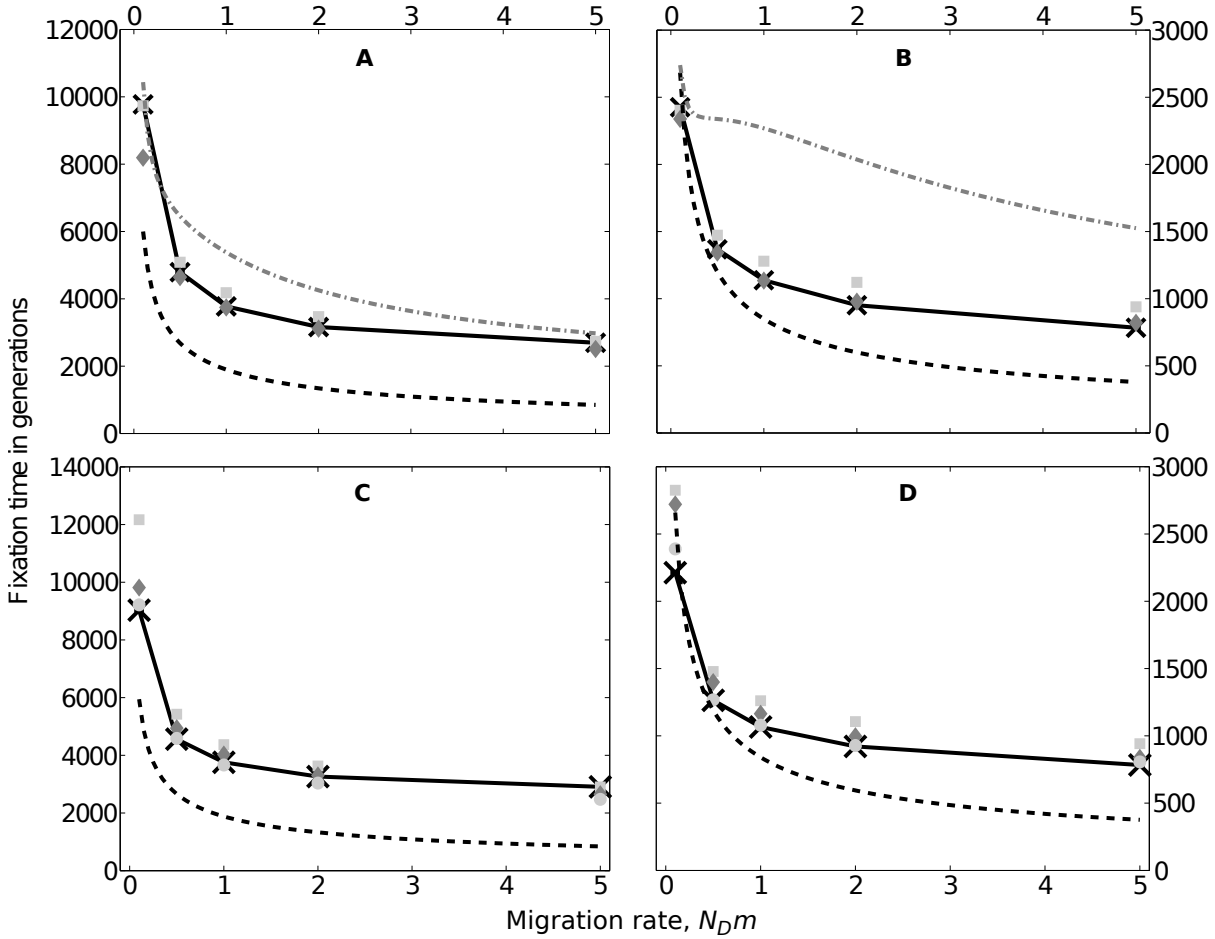


Figure 6.4: Fixation time of an advantageous allele as a function of the migration rate $N_D m$, where the population is divided over a one-dimensional structure with 11 demes ($D' = 6$; (a) and (b)), or a two-dimensional torus with 25 demes ($D' = 5$; (c) and (d)). $N_D = 2000$, and $N_D s = 10$ ((a) and (c)) or $N_D s = 50$ ((b) and (d)).

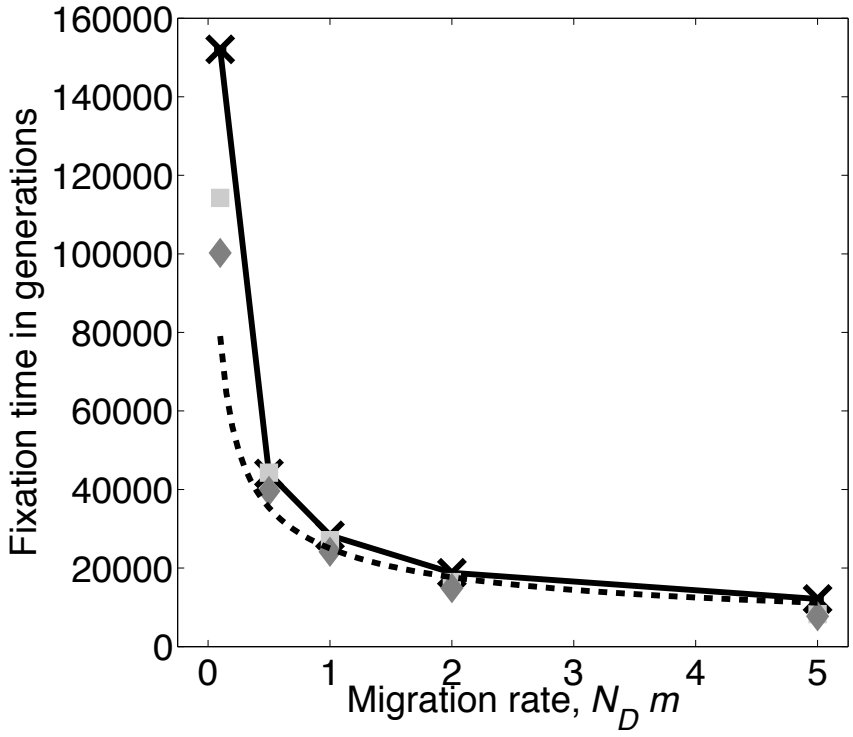


Figure 6.5: Fixation time of an advantageous allele, where the population is divided over a one-dimensional structure with 100 demes (so $D' = 50$), as a function of the migration rate. $N_D = 500$ and $Ns = 100$ (so that $N_D s = 1$).

except where migration is weak ($N_D m = 0.1$) in a one-dimensional population (Figures 6.2 and 6.3, (a) and (b); see also Figures 6.4 and 6.10). This second model needed to be corrected if applied to a two-dimensional structure with a large number of demes, due to the fact that an advantageous allele can take multiple routes between the deme that it first migrated to and to the most distant deme, with intermediate populations having different migration properties. This correction, when applied to the second model, makes it agree very accurately with simulation data. However, although the second model is generally more accurate for $N_D s = 10$, neither model appears to provide a consistent estimate with simulation data for this selective strength. This implies that for such weakly selected advantageous alleles, there are extra stochastic effects present that the model does not take into account, which should be investigated as part of future work in order to produce a more complete model. Model two also appear to be robust if there are a large number of demes (Figures 6.2 and 6.3, (c) and (d)), although results can be inaccurate in a two-dimensional model with weak selection and migration (see Figure 6.3(c), for example).

By analysing the model, a few key properties of mutant fixation become apparent. An advantageous allele can fix in a two-dimensional structure more quickly than in a one-dimensional model with the same number of demes. This is for two main reasons; first, it is clear that in the two-dimensional case, each deme is connected to more neighbours compared to a deme in a one-dimensional population, so a selected allele can spread through the entire population more quickly. This is reflected by the effective number of demes, D', across which an allele has to travel, which is greatly lower in two-dimensional structures ($D' = O(\sqrt{D})$, as opposed to $D' = O(D)$ in one-dimensional cases). A more original conclusion is that in a two-dimensional structure, the fixation time of an advantageous allele is greatly decreased due to the different paths that it can take. This means that some demes experience a greater input of migrants than others, so the allele will increase in frequency faster within such subpopulations. Therefore the

allele will spread faster overall. This conclusion is reflected in the correction applied to model two, which is needed in order to produce an accurate approximation for a large number of demes.

Generally this analysis has shown that the fixation time of an allele in a subdivided population is reduced by migration effects introducing more copies of an allele after it has established itself in a new deme. This behaviour may alter previously investigated effects of population subdivision, such as how levels of heterozygosity at linked neutral sites are changed, or whether there exists adequate gene flow between demes to prevent the populations from diverging. KIM and MARUKI (2011), for example, quantified and showed how the level of heterozygosity at a linked locus is greatly reduced in demes that lie nearest to where the sweep originated, reflecting how population subdivision delays the fixation of a novel advantageous allele, thus allowing more recombination to occur (see also BARTON (2000)). This analysis suggests that, since migration increases the speed at which the allele fixes in populations consisting of multiple demes, heterozygosity levels would not be broken down to a greater extent, compared to models where such migration effects were not considered. Future work should aim to implement the findings of this analysis into models of genetic hitchhiking, to accurately quantify how heterozygosity would be broken down in stepping-stone populations.

Secondly, this analysis can tells us more on whether gene flow among populations is so high that they cannot diverge, as discussed by EHRLICH and RAVEN (1969). In a review paper outlining existing data on migration rates, MORJAN and RIESEBERG (2004) found levels of gene flow to be higher than previously thought, but concluded that 'there are many species...that lack sufficient gene flow to prevent divergence'. This analysis demonstrates how even in populations with low migration levels, copies of new alleles can be transferred to new subpopulations by migration, thus increasing levels of gene flow between demes. The increase in fixation time can be substantial if subpopulations are closely connected, as in a two-dimensional model (see Figure 6.3, for example).

This analysis has also highlighted the need to investigate the manner in which populations are connected in natural systems, in order to understand how migration affects the spread of advantageous alleles. The degree of connectivity and type of population structure can have drastic effects on allele fixation time, so information on the manner in which communities are structured would also need to be estimated from field studies, in order to obtain an accurate estimate of fixation time.

Overall, this study highlights how even a modest amount of migration can affect the transfer of alleles into new demes, and decrease the fixation time of a selective sweep in structured populations. Future studies of hitchhiking, estimating the probability that neighbouring areas diverge, and other processes affected by population subdivision, should take this finding into account, in order in order to accurately determine the impact migration has on these.

6.A Derivation of migration coefficients used in models

In order to apply the models to different types of structured populations, the migration rate used in the models is not necessarily the same as the total migration rate between demes. Rather, the migration coefficient used in both models and Fisher's travelling-wave solution should be the variance in migration distance over demes, given the directions in which an individual can migrate (FISHER 1937).

For a one-dimensional model, denote a migration from a focal deme to the population to the left by -1, and a migration to the right by $+1$. Since it is equally likely that a migrant will move in either direction, the mean migration coefficient is $1/2{\cdot}(-1){+}1/2{\cdot}1 = 0$. Therefore the variance is $E[X^2]{-}E[X]^2 = (1/2{\cdot}1{+}1/2{\cdot}1){-}0 = 1$. So in one-dimensional populations, the migration rate used in both models and Fisher's solution is the same as the overall migration rate between demes.

This is easily extended to a two-dimensional torus model. Denote a migration from

a focal deme up to the population above it by the two-dimensional value $(0, 1)$; a migration to the right by $(1, 0)$; and the negative of these values for migrations down and left, respectively. Again since there is equal probability of a migration in either of these directions, the mean value is $(0, 0)$ and the variance is $(1/2, 1/2)$. Hence when applying model results to two-dimensional populations, the migration rate used is half that in simulations, as this is the variance between two adjacent demes (across the up-down or left-right axes). For Fisher's solution, the migration rate is scaled by the magnitude of this vector to represent diffusion across both dimensions. This is equal to $1/\sqrt{2}$.

6.B Supplementary Figures

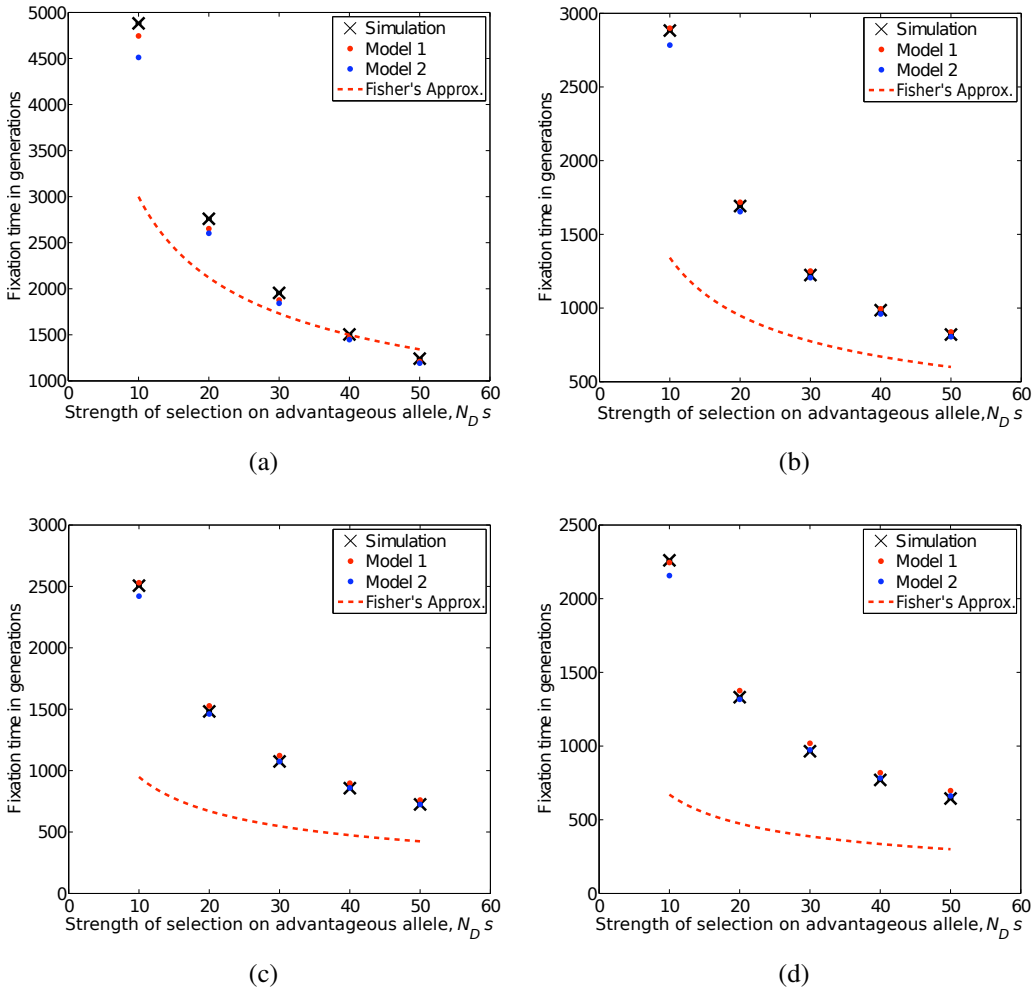


Figure 6.6: Fixation time of an advantageous allele where the population is divided over a one-dimensional structure with 5 demes ($D' = 3$). Results are plotted for the first model (red dots), second model (blue dots), simulation results (black crosses, standard errors lie within the markers) and Fisher's approximation (red dotted line). $N_D = 2000$, and (a) $N_D m = 0.1$, (b) $N_D m = 0.5$, (c) $N_D m = 1$ and (d) $N_D m = 2$.

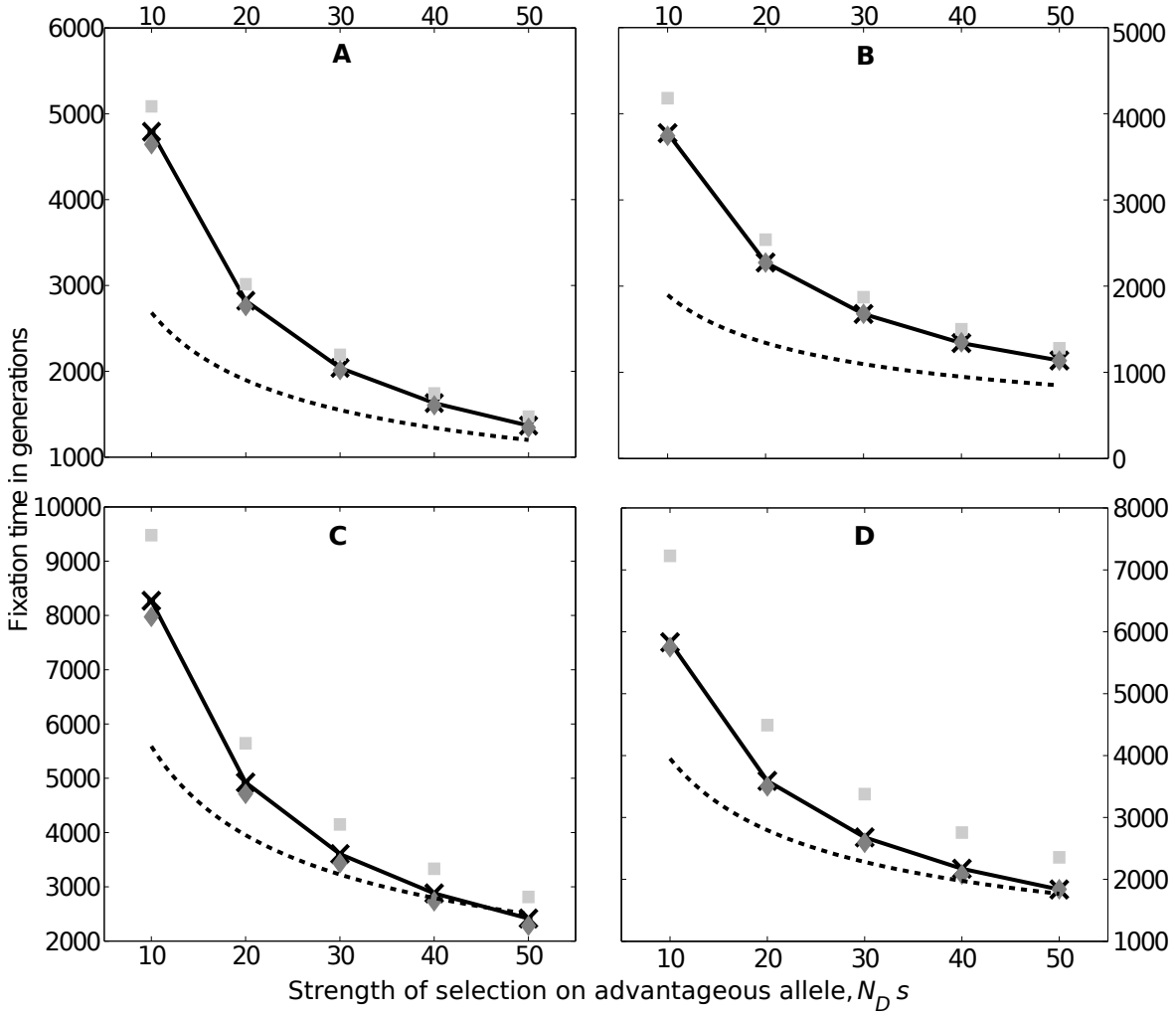


Figure 6.7: Fixation time of an advantageous allele where the population is divided over a one-dimensional structure with 11 demes ($D' = 6$; (a) and (b)), and with 101 demes ($D' = 50$; (c) and (d)). Results are plotted for the first model (light gray squares), second model (dark gray diamonds), simulation results (black crosses joined by a line, standard errors lie within the markers) and Fisher's approximation (black dotted line). $N_D = 2000$ ((a) and (b)) or $N = 500$ ((c) and (d)), and $N_D m = 0.5$ ((a) and (c)) or $N_D m = 1$ ((b) and (d)).

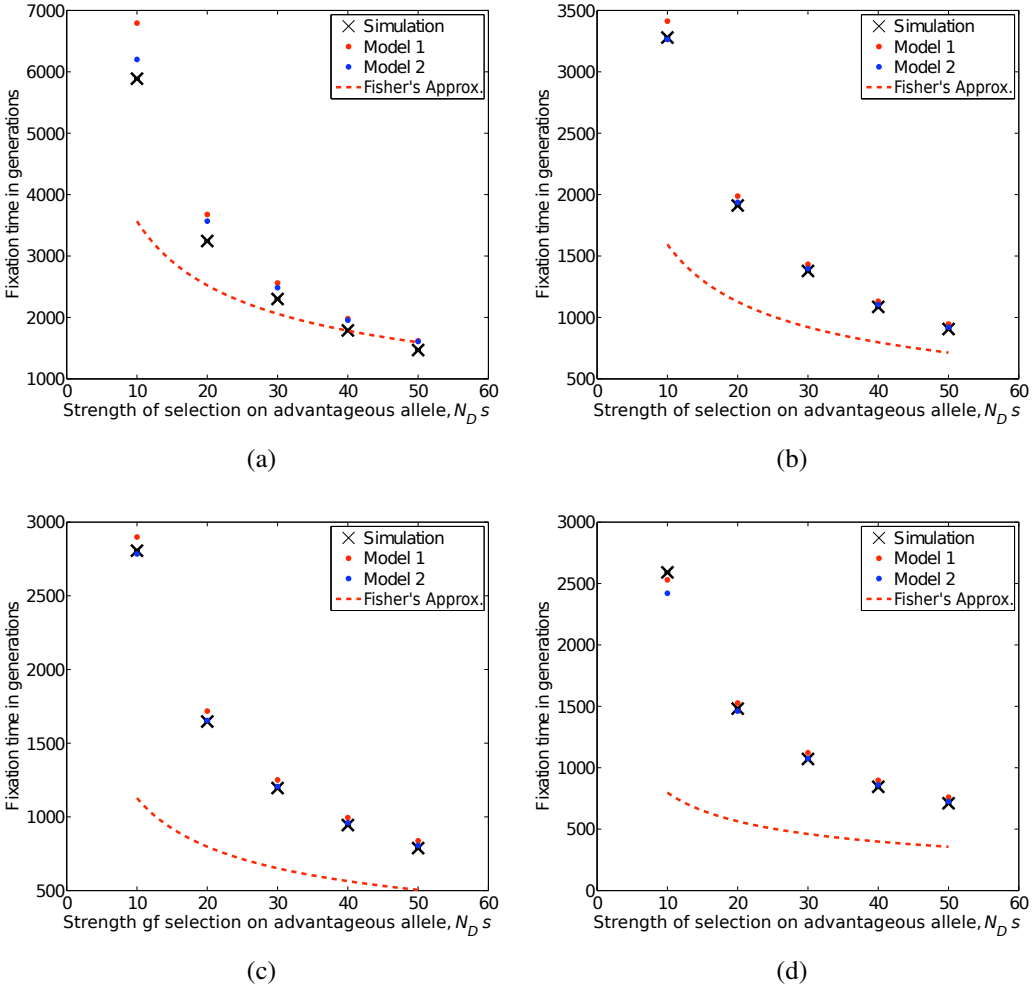


Figure 6.8: Fixation time of an advantageous allele where the population is divided over a two-dimensional structure with 9 demes (so $D' = 3$). $N_D = 2000$, and (a) $N_D m = 0.1$, (b) $N_D m = 0.5$, (c) $N_D m = 1$ and (d) $N_D m = 2$.

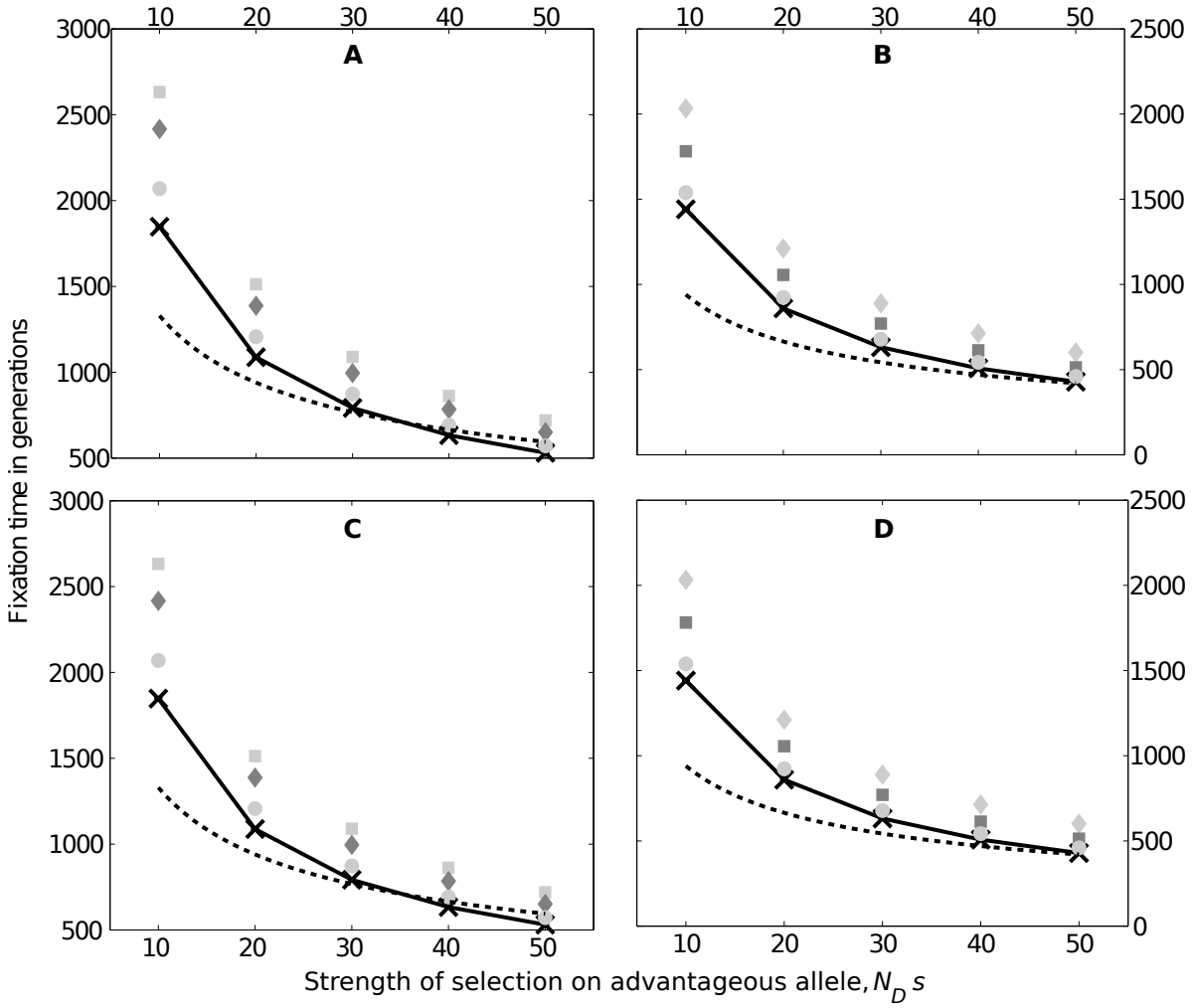


Figure 6.9: Fixation time of an advantageous allele where the population is divided over a two-dimensional structure with 25 demes ($D' = 5$; (a) and (b)), and 100 demes ($D' = 10$; (c) and (d)). As well as plotting the simulation data and two model results, the corrected version of model 2 that accounts for the different ways in which an advantageous allele can reach a target deme is also shown (light gray circles). $N_D = 2000$ ((a) and (b)) or $N = 500$ ((c) and (d)), and $N_D m = 0.5$ ((a) and (c)) or $N_D m = 1$ ((b) and (d)).

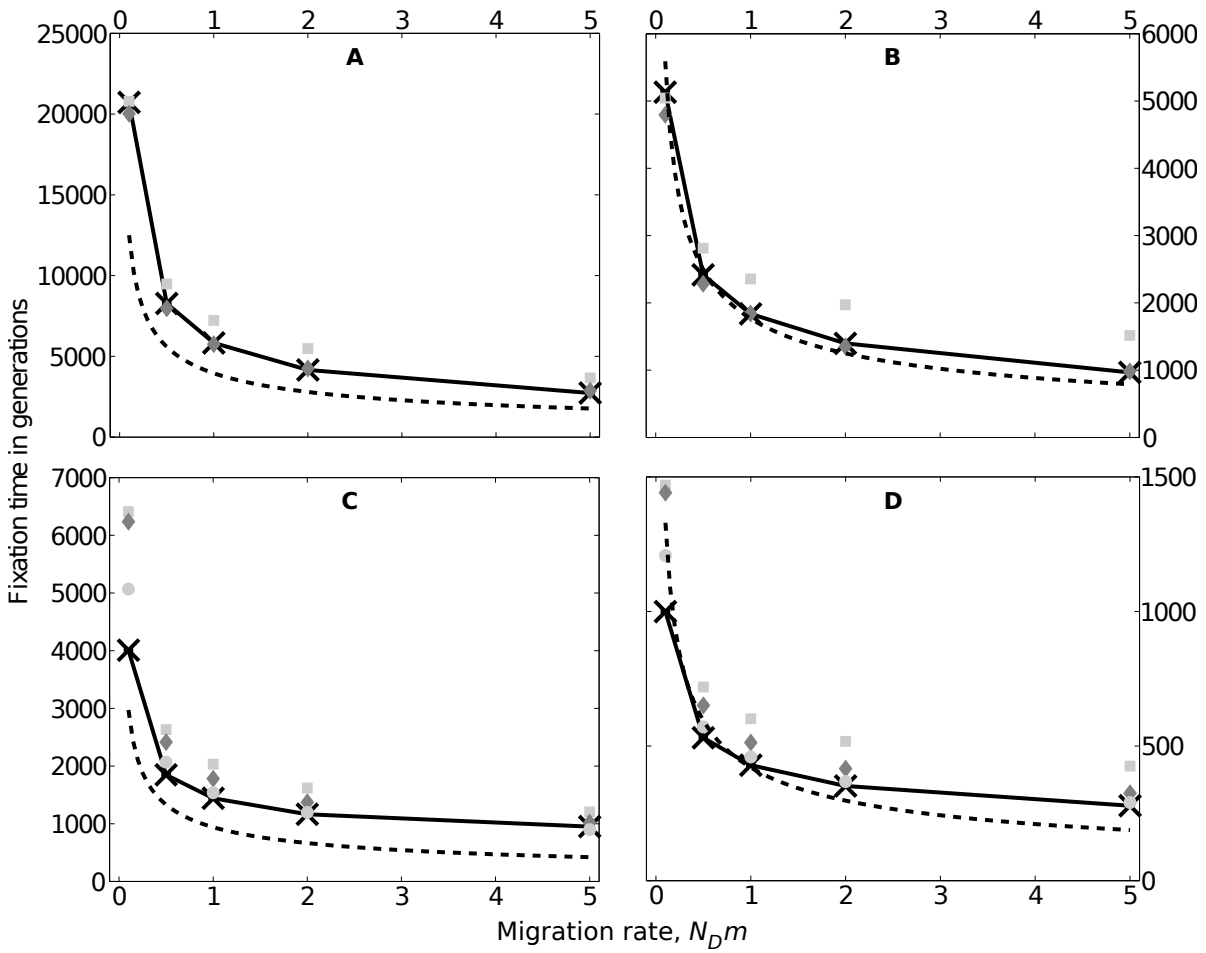
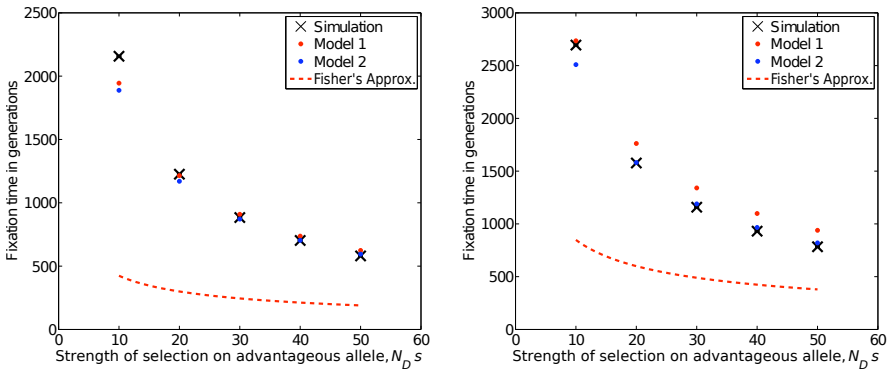
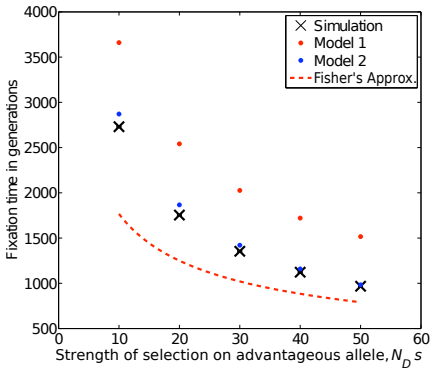


Figure 6.10: Fixation time of an advantageous allele as a function of the migration rate $N_D m$, where the population is divided over a one-dimensional structure with 101 demes ($D' = 50$; (a) and (b)), or a two-dimensional torus with 100 demes ($D' = 10$; (c) and (d)). $N_D = 500$, and $N_D s = 10$ ((a) and (c)) or $N_D s = 50$ ((b) and (d)).

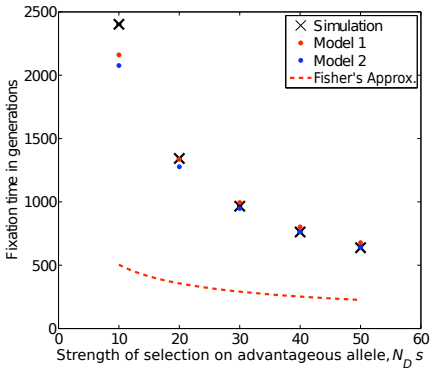


(a)

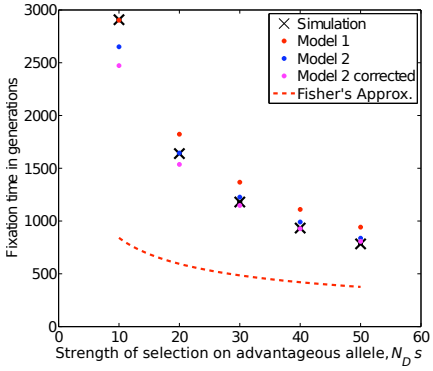
(b)



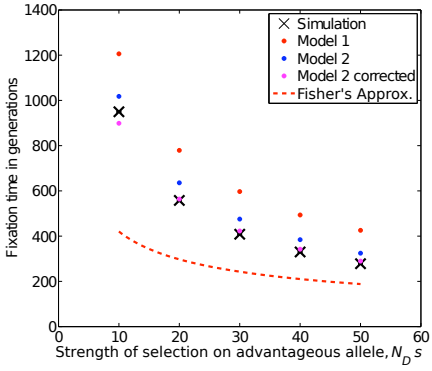
(c)



(d)



(e)



(f)

Figure 6.11: Fixation time of an advantageous allele for $N_D = 2000$ and $N_D m = 5$. The population is divided over a one-dimensional structure with 5 demes (a), 11 demes (b) or 101 demes (c), or a two-dimensional structure with 9 demes (d), 25 demes (e), or 100 demes (f).

Chapter 7

Conclusions and Discussion

7.1 Conclusions drawn from work undertaken for this thesis

Although there still does not exist a clear explanation as to why sex is ubiquitous, most of the fundamental processes that sex utilises are well-understood. The two most promising (and popular) hypotheses regarding the evolution of sex revolve around recombination breaking apart selection interference (BARTON 2010), and as a mechanism to produce rare genotypes that are less susceptible to co-evolving parasites (SALATHÉ *et al.* 2008a). The mutational-deterministic model has lost favour as a strong explanation, due to the lack of evidence for widespread synergistic epistasis in nature (KOUYOS *et al.* 2007). Other hypotheses, such as fitness-associated sex (HADANY and COMERON 2008) or the effect of heterogeneous environments (AGRAWAL 2009b) are intriguing, but require more thorough investigation to determine whether they can explain the widespread appearance of sex.

In this thesis, I aimed to further investigate the hypothesis that increased sex and recombination is selected for through breaking apart Hill-Robertson interference. Specifically, I aimed to extend the work of KEIGHTLEY and OTTO (2006), which demonstrated a large selective advantage for a recombination modifier if it acts over multiple loci subject to recurrent deleterious mutation. I first extended this work in Chapter 3, which demonstrated that the presence of advantageous mutation alongside deleterious mutation created significantly higher selection for increased recombination, compared to cases where mutation was just deleterious. However, selection acting on the modifier in the presence of both advantageous and deleterious mutation was sub-additive, in comparison to the individual advantageous-only and deleterious-only results. By analysing the variance in fitness present in a population over time, I showed that linkage disequilibrium was only slightly different in recombining populations, in comparison to asexuals. However, the genetic and genic variance was higher in the former scen-

ario (Figure 3.4), indicating that recombination was beneficial because it increased the response to selection. This reflects the advantage that recombination offers in moving novel advantageous mutants away from deleterious backgrounds, increasing their fixation probability (PECK 1994). It was also determined that the fixation probability of a recombination modifier did not increase linearly with higher mutation rates, and instead approached an asymptotic value once the mutation rate reached a high level (Figure 3.5(a)). This result arises as recombination breaks down linkage disequilibrium by the same magnitude, once the mutation rate became high enough (Figure 3.5(b)).

Chapter 4 outlined analytical methods that investigate how recombination disentangles selective sweeps from deleterious backgrounds, and prevent the fixation of deleterious alleles in sexuals. Both analytical and stochastic methods were developed in order to calculate this fixation probability (equations 4.5 and 4.24). These results were shown to be accurate in comparison to stochastic simulations (Figure 4.3), and can also be applied to the case where two linked advantageous alleles sweep to fixation (Figure 4.7). Furthermore, it was determined that the presence of linked deleterious alleles would rescue more neutral variation around the site of the sweep, compared to cases where deleterious hitchhikers were not present. This may make the sweep appear to be less strongly selected for than it actually is, based on the magnitude of neutral diversity in linked regions (Figure 4.2). A simple application of the model to human genetic data demonstrated that deleterious hitchhiking was unlikely to have occurred in recent evolutionary history, unless deleterious alleles were weakly selected against, or if selected loci were clustered together.

If the presence of recurrent advantageous and deleterious mutation across multiple loci selects for increased recombination, are these benefits large enough to overcome the costs of sex? This was investigated in Chapter 5, which tracked an asexual invading a subdivided sexual population. As the population size increased, lower levels of subdivision (measured using F_{st}) was needed to maintain sex. Even lower values were

required if the population was subject to advantageous and deleterious mutation, compared to deleterious-only scenarios. Despite these advantages to sex, the lowest level of population subdivision needed to maintain sex found was $F_{st} \approx 0.298$, for a population of $N = 40,000$ arranged in a one-dimensional structure, and both advantageous and deleterious mutation were present (Figure 5.2). This F_{st} value is generally high in comparison to those found in geographically-close populations (Appendix 5.A). The levels of F_{st} decreased if the number of demes increased (Figure 5.4), but higher levels of F_{st} was needed to maintain sex in two-dimensional populations (Figure 5.2), or if a small proportion of migrants were able to travel to any deme (Figure 5.3). It was found that the fixation probability of an invading asexual dropped approximately exponentially with a higher deleterious mutation rate, indicating a greater probability of asexuals experiencing mutational meltdown (Figure 5.5). However, a higher advantageous mutation rate had little effect on invading asexuals, if only advantageous mutations were present, indicating that they are ineffective at maintaining sex in structured populations (Figure 5.5). I also show that an asexual's fixation probability was minimised with intermediate values of s_d (strength of selection acting against deleterious mutations), which causes deleterious alleles to build-up at the fastest rate (Figure 5.6).

Finally, in order to further investigate the fate of asexuals spreading through a subdivided population, it needed to be determined how population structure affects the fixation time of a novel advantageous mutation. To this end, Chapter 6 outlines a general method to calculate the fixation time of a selective sweep in a subdivided population. The flexibility of the method means that it can be applied to different types of stepping-stone structures, and it was shown how the initial model could be adjusted to account for migration between multiple demes affecting the spread of an advantageous allele, as arises in a two-dimensional population. It was determined that the effect of migration transferring the allele between neighbouring demes had to be explicitly taken into account, so that models match with those from stochastic simulations. The method gen-

erally proved to be very accurate, although calculations overestimated the fixation time in two-dimensional populations with low migration rates ($N_D m \ll 1$ for N_D the size of the deme), and a large number of demes present (Figure 6.3).

7.2 Limitations of work undertaken in thesis

As with all scientific studies, the work undertaken in this thesis suffer from limitations due to specific assumptions used in the models. Ideally, these should be examined as part of future work, in order to determine whether breaking apart selection interference selects for sex and recombination under a wider range of biologically-realistic scenarios.

With regards to the multilocus simulations (Chapters 3 and 5), one limitation is that I used fixed values of the strengths of selection s_a, s_d, as well as fixing the proportion of mutations that are advantageous. Whilst I investigate how the dynamics of a recombination modifier (Figures 3.1), or an asexual invading a structured population (Figures 5.5, 5.12, 5.13) are altered as these values are changed, further analysis is needed to determine the impact that different values have on the level of F_{st} needed to maintain sex. Furthermore, the strength of deleterious mutants in nature would have a distribution of selective effects (EYRE-WALKER and KEIGHTLEY (2007)). Therefore, one key extension to these simulations would be to determine how strongly sex is selected for if the selective effect of mutations were chosen from a probability distribution, as opposed to having fixed effects.

A similar limitation is with regards to the maintenance of sex with different mutation rates. Most simulations in Chapter 5 assumed a deleterious mutation rate of $U = 1$. Whilst this is realistic for obligate sexuals, other species have lower mutation rates (BAER *et al.* 2007; HALLIGAN and KEIGHTLEY 2009). Figure 5.5 suggests that a greatly higher level of F_{st} is needed to maintain sex as the deleterious mutation rate decreases, but this needs to be formally shown. However, modest levels of F_{st} could

maintain facultative sexuals (that is, species that only reproduce a certain proportion of offspring sexually, and so have a lower overall cost of sex) if the deleterious mutation rate is reduced. The work in Chapter 5 could therefore be extended to consider what levels of population subdivision are needed to maintain facultative sexuals that are subject to lower mutation rates.

Individuals used in the simulations and analytical models were haploid for ease of computing. However, most obligate sexuals are diploid in nature, which can lead to differences in selection acting on sex and recombination. Specifically, if deleterious mutants are strongly recessive, then this can select against increased levels of sex and recombination, in populations subject to just deleterious mutation (ROZE 2009; ROZE and MICHOD 2010). Thus, the dynamics of sex in diploid populations should be investigated further, in order to gain a more complete picture of how sex is maintained in subdivided populations. If one is also investigating the effect of advantageous mutations on the evolution of sex in diploids, then dominance effects also needs to be considered. However, little is currently known about the value of dominance coefficients between advantageous mutations (ORR 2010).

With regards to the simulations in Chapter 5, it is assumed that selection occurs within each deme, and individuals only contribute to other regions due to migration. These models assumed complete local density regulation; that is, the total offspring produced within each deme is the same, and determined by the mean fitnss of that deme only. However, this is not reflective of all natural systems, since the size of each subpopulation can differ over generations, or demes can be subject to 'hard selection' (WALLACE 1975). The latter is defined as where the contribution of offspring from a deme to the entire population depends on the mean fitness of the individuals within it. These effects can again alter the fixation probability of asexuals, which should be investigated as part of future research. This can also affect the time to fixation of advantageous alleles, as determined in Chapter 6.

One of the main conclusions in Chapter 4 is that the region in which linked surrounding sites are likely to be fixed is reduced if deleterious alleles hitchhike with selective sweeps, compared to the case where only neutral alleles were linked to the sweep. However, it remains to be determined whether this result would still hold if a selective sweep arises whilst linked to several deleterious alleles. The overall outcome might be different because, with multiple loci, a novel selective sweep could be moved onto a more loaded background; this haplotype would not sweep to fixation as it is less fit than the original. Overall, this leads to fewer recombination events occurring over the course of the sweep, widening the average region in which linked sites are fixed, compared to cases where deleterious mutants were not present. However, this process would be very hard to investigate analytically, due to the numerous combinations of genetic backgrounds that arise via recombination. Therefore this problem could initially be tackled using multilocus simulations.

Finally, whilst the framework outlined in Chapter 6 measures the fixation time of a selective sweep in subdivided populations, this model should eventually be coupled with one of deleterious mutation accumulation, in order to determine how population subdivision increases the effect of Muller's Ratchet in asexuals. This more developed model can then be used to determine analytically how different types of population structure selects against asexuals, and whether costly sex can be maintained in populations that are currently too large to simulate.

7.3 Future directions for evolution of research on sex

Despite the major theoretical and empirical advances made in recent years, the question as to why costly sex is widespread is far from resolved. Whilst Hill-Robertson interference has gained substantial acceptance as the mechanism behind increased recombination (OTTO 2009; BARTON 2010), both selection-interference and Red Queen hypo-

theses provide cases where costly sex can be maintained (when acting over a subdivided population in the former case, and if generally strong parasitic selection is present in the latter case). Neither process, however, has yet to produce theoretical models that confidently account for the appearance of sex over a wide range of biological scenarios. This is especially true with regards to the evolution and maintenance of sex in diploids (ROZE 2009; ROZE and MICHOD 2010; AGRAWAL 2009a), despite the existence of empirical studies showing benefits to diploid sexuals (see, for example, MORRAN *et al.* (2009, 2011)). One of the major tasks facing theoreticians is in determining what scenarios leads to the maintenance of sex in diploids. Another major theoretical direction is investigating how multiple processes can aid selection for sex, such as the presence of both advantageous and deleterious mutation (Chapters 3, 5), or if hosts are subject both to parasitic infection and deleterious mutation accumulation in Red Queen models (HOWARD and LIVELY 1998, 2002). Such models generally show larger advantages to sex than if just one mechanism operates.

Theoretical work could also be extended to develop a more complete theory as to what conditions would favour facultative sex, as opposed to obligate sex. Some previous models showed that if a fraction of individuals reproduced sexually, the mean fitness of a population is increased to the same degree, compared to obligately sexual populations (PECK 1994; PECK *et al.* 1997; PECK and WAXMAN 2000). Similarly, models that demonstrate a benefit to obligate sex seldom show whether facultative sex can be maintained under a related mechanism, as opposed to evolving towards complete sexual reproduction. Developing a theory of the evolution of facultative sex could shed light on the initial appearance of sexual reproduction, and how it eventually evolved towards the appearance of obligate sex.

However, of greatest importance is the need to collect more experimental and empirical data, in order to test the large number of theoretical predictions. This situation has improved in recent years, with numerous studies showing that sex can be advant-

ageous through breaking apart selection interference (COLEGRAVE 2002; GODDARD *et al.* 2005), by co-evolving with parasites (MORRAN *et al.* 2011), when individuals are subject to a stressed environment (SCHOUSTRA *et al.* 2010), and by switching between heterogeneous environments (BECKS and AGRAWAL 2010). However, more experimental evidence is needed to discriminate between the major hypotheses, as well as more field evidence to determine what mechanism selects for sex in nature. Therefore, future experiments should try not just to determine what environments would select for sex, but also attempt to separate the effects of directional selection from fluctuating environments. This information can then be used to inform whether sex is selected for in fluctuating conditions due to the presence of homogeneous environments or parasite interactions, or whether it instead evolves as a means to break apart selection interference and increase the response to selection.

More data are also needed in order to test and verify specific predictions made by theoretical work, in light of recent theoretical predictions. Specifically, more studies similar to that of ZEYL and BELL (1997) are needed to determine whether the presence of advantageous mutations increase selection for sex and recombination, or whether the main advantages lie in purging deleterious alleles. All major theories for the evolution of sex, as well as newly-emerging ones, postulate that sex is selected for in order to improve adaptation. Examples of this idea include sex increasing fixation probabilities of advantageous alleles; increasing individuals' resistance to parasitic infection; or increasing rate of adaptation to heterogeneous environments. Based on this idea, future experiments should not only test the role that adaptive mutations play in selecting for sex, but also determine what proportion of mutation needs to be advantageous to ensure that sex is beneficial. However, theory outlined in this thesis suggests that the strength of both beneficial and deleterious alleles, as well as the deleterious mutation rate, is also important. Therefore, experimental studies also need to determine these mutational parameters, so that experiments can be fully related to the existing theory. This is clearly

an onerous task, and so it remains to be seen whether these experiments can actually be run in practice.

It also needs to be verified whether homogeneous spatial populations cause a sufficient mutation meltdown in asexuals that can lead to the maintenance of sex. More empirical studies should also determine whether the evolution of sex and recombination leads to an increase in fitness variance, as this observation can inform on the underlying mechanism (BECKS and AGRAWAL 2011).

Whilst there has been considerable success in verifying predictions made by the Red Queen hypothesis, recent theoretical advances have yet to be fully investigated. These include whether weakly virulent parasites select for sex (LIVELY 2009, 2010b), or the effect of multiple parasites on the maintenance of sexuals (MOSTOWY *et al.* 2010). The outcome of these studies could provide key insights into which genetic processes are the main determinant in the evolution of sex, and might even provide an answer to this notorious scientific enigma.

7.4 Overall Summary

In summary, the work in this thesis strengthens the case that breaking apart selection interference across multiple loci explains widespread costly sex, due to the presence of recurrent advantageous mutation. Recombination can be strongly selected for by breaking apart selection interference in the presence of both recurrent advantageous and deleterious mutation, as this stops the buildup of deleterious mutants in the population, as well as improving the fixation probability of adaptive mutations. This means that finite-population models can now explain selection for increased recombination due to adaptive mutation, without relying on heterogeneous-environment or fitness-associated mechanisms. These benefits to recombination can maintain costly sex in subdivided populations with lower levels of F_{st}, compared to cases where mutation is just deleteri-

ous. However, the lowest F_{st} values observed in simulations are higher than those found in many natural studies over a small geographic range, but critical values could potentially be greatly lower in very large populations that are spread over many demes. Overall, since recurrent advantageous and deleterious mutations are ubiquitous in nature, and all populations are subdivided to some degree, it is possible that all three of these mechanisms could act together to prevent asexuals displacing sexuals, leading to the widespread prevalence of sex in large populations.

Bibliography

ABDULLAH, M. F. F., and R. H. BORTS, 2001 Meiotic recombination frequencies are affected by nutritional states in *Saccharomyces cerevisiae*. Proc. Natl. Acad. Sci. USA **98**: 14524–14529.

ABRAMOWITZ, M., and I. STEGUN, 1970 *Handbook of Mathematical Functions*. Dover Publications, Inc., New York.

AGRAWAL, A. F., 2001 Sexual selection and the maintenance of sexual reproduction. Nature **411**: 692–695.

AGRAWAL, A. F., 2006 Similarity selection and the evolution of sex: Revisiting the Red Queen. PLoS Biol. **4**: e265.

AGRAWAL, A. F., 2009a Differences between selection on sex verses recombination in Red Queen models with diploid hosts. Evolution **63**: 2131–2141.

AGRAWAL, A. F., 2009b Spatial heterogeneity and the evolution of sex in diploids. Am. Nat. **174**: S54–S70.

AGRAWAL, A. F., and J. R. CHASNOV, 2001 Recessive mutations and the maintenance of sex in structured populations. Genetics **158**: 913–917.

AGRAWAL, A. F., L. HADANY, and S. P. OTTO, 2005 The evolution of plastic recombination. Genetics **171**: 803–812.

AGUINAGALDE, I., A. HAMPE, A. MOHANTY, J. P. MARTÍN, J. DUMINIL, *et al.*, 2005 Effects of life-history traits and species distribution on genetic structure at maternally inherited markers in European trees and shrubs. J. Biogeogr. **32**: 329–339.

ALBERTO, F., P. T. RAIMONDI, D. C. REED, N. C. COELHO, R. LEBLOIS, *et al.*, 2010 Habitat continuity and geographic distance predict population genetic differentiation in giant kelp. Ecology **91**: 49–56.

ANDOLFATTO, P., 2005 Adaptive evolution of non-coding DNA in *Drosophila*. Nature **437**: 1149–1152.

ANDOLFATTO, P., 2007 Hitchhiking effects of recurrent beneficial amino acid substitutions in the *Drosophila melanogaster* genome. Genome Res. **17**: 1755–1762.

BACHTROG, D., and I. GORDO, 2004 Adaptive evolution of asexual populations under Muller's ratchet. Evolution **58**: 1403–1413.

BAER, C. F., M. M. MIYAMOTO, and D. R. DENVER, 2007 Mutation rate variation in multicellular eukaryotes: causes and consequences. Nat. Rev. Genet. **8**: 619–631.

BARRETT, R. D. H., R. CRAIG MACLEAN, and G. BELL, 2006 Mutations of intermediate effect are responsible for adaptation in evolving *Pseudomonas fluorescens* populations. Biol. Lett. **2**: 236–238.

BARTON, N. H., 1993 The probability of fixation of a favoured allele in a subdivided population. Genet. Res. **62**: 149–157.

BARTON, N. H., 1994 The reduction in fixation probability caused by substitutions at linked loci. Genet. Res. **64**: 199–208.

BARTON, N. H., 1995a A general model for the evolution of recombination. Genet. Res. **65**: 123–144.

BARTON, N. H., 1995b Linkage and the limits to natural selection. Genetics **140**: 821–841.

BARTON, N. H., 2000 Genetic hitchhiking. Philos. Trans. R. Soc. Lond. B Biol. Sci. **355**: 1553–1562.

BARTON, N. H., 2009 Why sex and recombination? Cold Spring Harb. Symp. Quant. Biol. .

BARTON, N. H., 2010 Genetic linkage and natural selection. Philos. Trans. R. Soc. Lond. B Biol. Sci. **365**: 2559–2569.

BARTON, N. H., and B. CHARLESWORTH, 1998 Why sex and recombination? Science **281**: 1986–1990.

BARTON, N. H., and S. P. OTTO, 2005 Evolution of recombination due to random drift. Genetics **169**: 2353–2370.

BARTON, N. H., and M. TURELLI, 1991 Natural and sexual selection on many loci. Genetics **127**: 229–255.

BAUDRY, E., B. VIGINIER, and M. VEUILLE, 2004 Non-African populations of *Drosophila melanogaster* have a unique origin. Mol. Biol. Evol. **21**: 1482–1491.

BECKS, L., and A. F. AGRAWAL, 2010 Higher rates of sex evolve in spatially heterogeneous environments. Nature **468**: 89–92.

BECKS, L., and A. F. AGRAWAL, 2011 The effect of sex on the mean and variance of fitness in facultatively sexual rotifers. J. Evol. Biol. **24**: 656–664.

BERNSTEIN, H., F. A. HOPF, and R. E. MICHOD, 1988 Is Meiotic Recombination an Adaptation for repairing DNA, Producing Genetic Variation, or Both? In R. E.

Michod and B. R. Levin, editors, *The Evolution of Sex*. Sinauer Press, Sunderland, Massachusetts, 139–160.

BETANCOURT, A. J., and D. C. PRESGRAVES, 2002 Linkage limits the power of natural selection in *Drosophila*. Proc. Natl. Acad. Sci. USA **99**: 13616–13620.

BETANCOURT, A. J., J. J. WELCH, and B. CHARLESWORTH, 2009 Reduced effectiveness of selection caused by a lack of recombination. Curr. Biol. **19**: 655–660.

BIERNE, N., 2010 The distinctive footprints of local hitchhiking in a varied environment and global hitchhiking in a subdivided population. Evolution **64**: 3254–3272.

BIERNE, N., and A. EYRE-WALKER, 2004 The genomic rate of adaptive amino acid substitution in *Drosophila*. Mol. Biol. Evol. **21**: 1350–1360.

BLACHFORD, A., and M. DOEBELI, 2009 On luck and sex. Evolution **63**: 40–47.

BOYKO, A. R., S. H. WILLIAMSON, A. R. INDAP, J. D. DEGENHARDT, R. D. HERNANDEZ, *et al.*, 2008 Assessing the evolutionary impact of amino acid mutations in the human genome. PLoS Genet. **4**: e1000083.

BROMAN, K. W., J. C. MURRAY, V. C. SHEFFIELD, R. L. WHITE, and J. L. WEBER, 1998 Comprehensive human genetic maps: Individual and sex-specific variation in recombination. Am. J. Hum. Genet. **63**: 861 – 869.

BULL, J. J., M. R. BADGETT, and H. A. WICHMAN, 2000 Big-benefit mutations in a bacteriophage inhibited with heat. Mol. Biol. Evol. **17**: 942–950.

BULMER, M. G., 1976 The effect of selection on genetic variability: a simulation study. Genet. Res. **28**: 101–117.

BULMER, M. G., 1980 *The mathematical theory of quantitative genetics*. Clarendon Press, Oxford.

BURT, A., 2000 Sex, recombination, and the efficacy of selection - was Weismann right? Evolution **54**: 337–351.

CHARLESWORTH, B., 1976 Recombination modification in a fluctuating environment. Genetics **83**: 181–195.

CHARLESWORTH, B., 1990 Mutation-selection balance and the evolutionary advantage of sex and recombination. Genet. Res. **55**: 199–221.

CHARLESWORTH, B., 1994 The effect of background selection against deleterious mutations on weakly selected, linked variants. Genet. Res. **63**: 213–227.

CHARLESWORTH, B., A. J. BETANCOURT, V. B. KAISER, and I. GORDO, 2009 Genetic recombination and molecular evolution. Cold Spring Harb. Symp. Quant. Biol. **74**: 177–186.

CHARLESWORTH, B., and D. CHARLESWORTH, 1975 An experiment on recombination load in *Drosophila melanogaster*. Genet. Res. **25**: 267–274.

CHARLESWORTH, B., and D. CHARLESWORTH, 1997 Rapid fixation of deleterious alleles can be caused by muller's ratchet. Genet. Res. **70**: 63–73.

CHARLESWORTH, B., M. T. MORGAN, and D. CHARLESWORTH, 1993 The effect of deleterious mutations on neutral molecular variation. Genetics **134**: 1289–1303.

CHARLESWORTH, D., B. CHARLESWORTH, and M. T. MORGAN, 1995 The pattern of neutral molecular variation under the background selection model. Genetics **141**: 1619–1632.

CHARLESWORTH, D., B. CHARLESWORTH, and C. STROBECK, 1977 Effects of selfing on selection for recombination. Genetics **86**: 213–226.

CHASNOV, J. R., 2000 Mutation-selection balance, dominance and the maintenance of sex. Genetics **156**: 1419–1425.

CHERRY, J. L., 2003a Selection in a subdivided population with dominance or local frequency dependence. Genetics **163**: 1511–1518.

CHERRY, J. L., 2003b Selection in a subdivided population with local extinction and recolonization. Genetics **164**: 789–795.

CHERRY, J. L., 2004 Selection, subdivision and extinction and recolonization. Genetics **166**: 1105–1114.

CHRISTIANSEN, F. B., S. P. OTTO, A. BERGMAN, and M. W. FELDMAN, 1998 Waiting with and without recombination: The time to production of a double mutant. Theor. Popul. Biol. **53**: 199–215.

CHUN, S., and J. C. FAY, 2011 Evidence for hitchhiking of deleterious mutations within the human genome. PLoS Genet. **7**: e1002240.

CLEGG, M. T., J. F. KIDWELL, and C. R. HORCH, 1980 Dynamics of correlated genetic systems. V. Rates of decay of linkage disequilibria in experimental populations of *Drosophila melanogaster*. Genetics **94**: 217–234.

COLEGRAVE, N., 2002 Sex releases the speed limit on evolution. Nature **420**: 664–666.

COMERON, J. M., A. WILLIFORD, and R. M. KLIMAN, 2008 The Hill-Robertson effect: evolutionary consequences of weak selection and linkage in finite populations. Heredity **100**: 19–31.

COOPER, T. F., 2007 Recombination speeds adaptation by reducing competition between beneficial mutations in populations of *Escherichia coli*. PLoS Biol. **5**: e225.

CROW, J. F., 1970 Genetic loads and the cost of natural selection. In K.-I. Kojima, editor, *Mathematical Topics in Population Genetics*, volume 1 of *Biomathematics*. Springer-Verlag, Berlin, 128—177.

DENVER, D. R., K. MORRIS, M. LYNCH, and W. K. THOMAS, 2004 High mutation rate and predominance of insertions in the *Caenorhabditis elegans* nuclear genome. Nature **430**: 679–682.

DESAI, M. M., and D. S. FISHER, 2007 Beneficial mutation-selection balance and the effect of linkage on positive selection. Genetics **176**: 1759–1798.

DIERINGER, D., V. NOLTE, and C. SCHLOTTERER, 2005 Population structure in African *Drosophila melanogaster* revealed by microsatellite analysis. Mol. Ecol. **14**: 563–573.

DONCASTER, C. P., G. E. POUND, and S. J. COX, 2000 The ecological cost of sex. Nature **404**: 281–285.

DUMINIL, J., S. FINESCHI, A. HAMPE, P. JORDANO, D. SALVINI, *et al.*, 2007 Can Population Genetic Structure Be Predicted from Life–History Traits? Am. Nat. **169**: 662–672.

DYBDAHL, M. F., and C. M. LIVELY, 1995 Diverse, endemic and polyphyletic clones in mixed populations of a freshwater snail (*Potamopyrgus antipodarum*). J. Evol. Biol. **8**: 385–398.

EHRLICH, P. R., and P. H. RAVEN, 1969 Differentiation of populations. Science **165**: 1228–1232.

ELENA, S. F., and R. E. LENSKI, 2003 Evolution experiments with microorganisms: the dynamics and genetic bases of adaptation. Nat. Rev. Genet. **4**: 457–469.

ENGELMOER, D. J. P., and D. E. ROZEN, 2011 Competence increases survival during stress in *Streptococcus pneumoniae*. Evolution **65**: 3475–3485.

EÖRY, L., D. L. HALLIGAN, and P. D. KEIGHTLEY, 2010 Distributions of selectively constrained sites and deleterious mutation rates in the hominid and murid genomes. Mol. Biol. Evol. **27**: 177–192.

EWENS, W. J., 2004 *Mathematical population genetics : 1. Theoretical introduction*, volume 27 of *Interdisciplinary applied mathematics*. Springer, New York, second edition.

EYRE-WALKER, A., 2006 The genomic rate of adaptive evolution. Trends Ecol. Evol. **21**: 569–575.

EYRE-WALKER, A., and P. D. KEIGHTLEY, 2007 The distribution of fitness effects of new mutations. Nat. Rev. Genet. **8**: 610–618.

EYRE-WALKER, A., and P. D. KEIGHTLEY, 2009 Estimating the rate of adaptive molecular evolution in the presence of slightly deleterious mutations and population size change. Mol. Biol. Evol. **26**: 2097–2108.

EYRE-WALKER, A., M. WOOLFIT, and T. PHELPS, 2006 The distribution of fitness effects of new deleterious amino acid mutations in humans. Genetics **173**: 891–900.

FELDMAN, M. W., F. B. CHRISTIANSEN, and L. D. BROOKS, 1980 Evolution of recombination in a constant environment. Proc. Natl. Acad. Sci. USA **77**: 4838–4841.

FELSENSTEIN, J., 1974 The evolutionary advantage of recombination. Genetics **78**: 737–756.

FISCHER, O., and P. SCHMID-HEMPEL, 2005 Selection by parasites may increase host recombination frequency. Biol. Lett. **1**: 193–195.

FISHER, R. A., 1930 *The genetical theory of natural selection*. The Clarendon Press, Oxford.

FISHER, R. A., 1937 The wave of advance of advantageous genes. Ann. Eugen. **7**: 355–369.

FREVILLE, H., F. JUSTY, and I. OLIVIERI, 2001 Comparative allozyme and microsatellite population structure in a narrow endemic plant species, *Centaurea corymbosa* Pourret (Asteraceae). Mol. Ecol. **10**: 879–889.

GABRIEL, W., M. LYNCH, and R. BURGER, 1993 Muller's ratchet and mutational meltdowns. Evolution **47**: 1744–1757.

GANDON, S., and S. P. OTTO, 2007 The evolution of sex and recombination in response to abiotic or coevolutionary fluctuations in epistasis. Genetics **175**: 1835–1853.

GARCÍA-DORADO, A., C. LÓPEZ-FANJUL, and A. CABALLERO, 1999 Properties of spontaneous mutations affecting quantitative traits. Genet. Res. **74**: 341–350.

GERRISH, P., and R. LENSKI, 1998 The fate of competing beneficial mutations in an asexual population. Genetica **102–103**: 127–144.

GESSLER, D. D. G., 1995 The constraints of finite size in asexual populations and the rate of the ratchet. Genet. Res. **66**: 241–253.

GESSLER, D. D. G., and S. XU, 2000 Meiosis and the evolution of recombination at low mutation rates. Genetics **156**: 449–456.

GODDARD, M. R., H. C. J. GODFRAY, and A. BURT, 2005 Sex increases the efficacy of natural selection in experimental yeast populations. Nature **434**: 636–640.

GODDARD, M. R., D. GREIG, and A. BURT, 2001 Outcrossed sex allows a selfish gene to invade yeast populations. Proc. R. Soc. B **268**: 2537–2542.

GORDO, I., and P. CAMPOS, 2006 Adaptive evolution in a spatially structured asexual population. Genetica **127**: 217–229.

GORDO, I., and P. R. A. CAMPOS, 2008 Sex and deleterious mutations. Genetics **179**: 621–626.

GORDO, I., and B. CHARLESWORTH, 2000 The degeneration of asexual haploid populations and the speed of Muller's ratchet. Genetics **154**: 1379–1387.

GRISHKAN, I., A. B. KOROL, E. NEVO, and S. P. WASSER, 2003 Ecological stress and sex evolution in soil microfungi. Proc. R. Soc. B **270**: 13–18.

HAAG, C. R., and D. ROZE, 2007 Genetic load in sexual and asexual diploids: Segregation, dominance and genetic drift. Genetics **176**: 1663–1678.

HAAG-LIAUTARD, C., M. DORRIS, X. MASIDE, S. MACASKILL, D. L. HALLIGAN, *et al.*, 2007 Direct estimation of per nucleotide and genomic deleterious mutation rates in *Drosophila*. Nature **445**: 82–85.

HADANY, L., and T. BEKER, 2003 On the evolutionary advantage of fitness-associated recombination. Genetics **165**: 2167–2179.

HADANY, L., and J. M. COMERON, 2008 Why are sex and recombination so common? Ann. NY Acad. Sci. **1133**: 26–43.

HADANY, L., and M. W. FELDMAN, 2005 Evolutionary traction: the cost of adaptation and the evolution of sex. J. Evol. Biol. **18**: 309–314.

HADANY, L., and S. P. OTTO, 2007 The evolution of condition-dependent sex in the face of high costs. Genetics **176**: 1713–1727.

HADANY, L., and S. P. OTTO, 2009 ConditionâĂŘdependent sex and the rate of adaptation. Am. Nat. **174**: S71–S78.

HALDANE, J. B. S., 1924 A mathematical theory of natural and artificial selection, part I. Trans. Cambridge Philos. Soc. **23**: 19–41.

HALDANE, J. B. S., 1927 A mathematical theory of natural and artificial selection, part V: Selection and mutation. Math. Proc. Cambridge Philos. Soc. **23**: 838–844.

HALL, D. W., and S. B. JOSEPH, 2010 A high frequency of beneficial mutations across multiple fitness components in *Saccharomyces cerevisiae*. Genetics **185**: 1397–1409.

HALLIGAN, D. L., and P. D. KEIGHTLEY, 2006 Ubiquitous selective constraints in the *Drosophila* genome revealed by a genome-wide interspecies comparison. Genome Res. **16**: 875–884.

HALLIGAN, D. L., and P. D. KEIGHTLEY, 2009 Spontaneous mutation accumulation studies in evolutionary genetics. Annu. Rev. Ecol. Systematics **40**: 151–172.

HALLIGAN, D. L., F. OLIVER, A. EYRE-WALKER, B. HARR, and P. D. KEIGHTLEY, 2010 Evidence for pervasive adaptive protein evolution in wild mice. PLoS Genet. **6**: e1000825.

HAMILTON, W. D., 1980 Sex versus non-sex versus parasite. Oikos **35**: 282–290.

HAMILTON, W. D., R. AXELROD, and R. TANESE, 1990 Sexual reproduction as an adaptation to resist parasites (a review). Proc. Natl. Acad. Sci. USA **87**: 3566–3573.

HAMRICK, J., 1989 Isozymes and the Analysis of Genetic Structure in Plant Populations. In D. E. Soltis and P. S. Soltis, editors, *Isozymes in plant biology*. Dioscorides Press, Portland, Or., 87–105.

HARTFIELD, M., and S. P. OTTO, 2011 Recombination and hitchhiking of deleterious alleles. Evolution **65**: 2421–2434.

HARTFIELD, M., S. P. OTTO, and P. D. KEIGHTLEY, 2010 The role of advantageous mutations in enhancing the evolution of a recombination modifier. Genetics **184**: 1153–1164.

HARTFIELD, M., S. P. OTTO, and P. D. KEIGHTLEY, 2012 Can realistic population subdivision maintain sex in finite, structured populations? In prep.

HARTL, D. L., and A. G. CLARK, 2007 *Principles of population genetics*. Sinauer Associates, Sunderland, Mass., 4th edition.

HEDRICK, P. W., 1983 *Genetics of populations*. Jones and Bartlett, Boston.

HEDRICK, P. W., 2005 A standardized genetic differentiation measure. Evolution **59**: 1633–1638.

HELLER, R., and H. R. SIEGISMUND, 2009 Relationship between three measures of genetic differentiation G_{st}, D_{est} and G'_{ST}: how wrong have we been? Mol. Ecol. **18**: 2080–2083.

HICKEY, D. A., 1982 Selfish DNA: A sexually-transmitted nuclear parasite. Genetics **101**: 519–531.

HIGGINS, K., and M. LYNCH, 2001 Metapopulation extinction caused by mutation accumulation. Proc. Natl. Acad. Sci. USA **98**: 2928–2933.

HILL, J. A., and S. P. OTTO, 2007 The role of pleiotropy in the maintenance of sex in yeast. Genetics **175**: 1419–1427.

HILL, W. G., and A. ROBERTSON, 1966 The effect of linkage on limits to artificial selection. Genet. Res. **8**: 269–294.

HOWARD, R. S., and C. M. LIVELY, 1998 The maintenance of sex by parasitism and mutation accumulation under epistatic fitness functions. Evolution **52**: 604–610.

HOWARD, R. S., and C. M. LIVELY, 2002 The ratchet and the Red Queen: the maintenance of sex in parasites. J. Evol. Biol. **15**: 648–656.

HUDSON, R. R., and N. L. KAPLAN, 1995 Deleterious background selection with recombination. Genetics **141**: 1605–1617.

HUDSON, R. R., M. KREITMAN, and M. AGUADE, 1987 A test of neutral molecular evolution based on nucleotide data. Genetics **116**: 153–159.

HUGHES, J. M., D. J. SCHMIDT, A. MCLEAN, and A. WHEATLEY, 2008 Population genetic structure in stream insects: What have we learned? In J. Lancaster and R. A. Briers, editors, *Aquatic insects: challenges to populations: proceedings of the Royal Entomological Society's 24th symposium*, chapter 14. CABI, Wallingford, Oxfordshire ; Cambridge, Mass., 268–288.

ILES, M. M., K. WALTERS, and C. CANNINGS, 2003 Recombination can evolve in large finite populations given selection on sufficient loci. Genetics **165**: 2249–2258.

JAENIKE, J., 1978 An hypothesis to account for the maintenance of sex within populations. Evol. Theory **3**: 191–194.

JENSEN, J. D., K. R. THORNTON, and P. ANDOLFATTO, 2008 An approximate Bayesian estimator suggests strong, recurrent selective sweeps in *Drosophila*. PLoS Genet. **4**: e1000198.

JOHNSON, T., and N. H. BARTON, 2002 The effect of deleterious alleles on adaptation in asexual populations. Genetics **162**: 395–411.

JOKELA, J., M. F. DYBDAHL, and C. M. LIVELY, 2009 The maintenance of sex, clonal dynamics, and host-parasite coevolution in a mixed population of sexual and asexual snails. Am. Nat. **174**: S43–S53.

JOKELA, J., C. M. LIVELY, M. F. DYBDAHL, and J. A. FOX, 1997 Evidence for a cost of sex in the freshwater snail *Potamopyrgus antipodarum*. Ecology **78**: 452–460.

JOKELA, J., C. M. LIVELY, M. F. DYBDAHL, and J. A. FOX, 2003 Genetic variation in sexual and clonal lineages of a freshwater snail. Biol. J. Linn. Soc. **79**: 165–181.

JORDE, L. B., M. BAMSHAD, and A. R. ROGERS, 1998 Using mitochondrial and nuclear DNA markers to reconstruct human evolution. BioEssays **20**: 126–136.

JOSEPH, S. B., and D. W. HALL, 2004 Spontaneous mutations in diploid *Saccharomyces cerevisiae*: More beneficial than expected. Genetics **168**: 1817–1825.

JOST, L., 2008 G_{st} and its relatives do not measure differentiation. Mol. Ecol. **17**: 4015–4026.

KAISER, V. B., and B. CHARLESWORTH, 2009 The effects of deleterious mutations on evolution in non-recombining genomes. Trends Genet. **25**: 9–12.

KARASOV, T., P. W. MESSER, and D. A. PETROV, 2010 Evidence that adaptation in *Drosophila* is not limited by mutation at single sites. PLoS Genet. **6**: e1000924.

KARLIN, E. F., S. C. HOTCHKISS, S. B. BOLES, H. K. STENØIEN, K. HASSEL, *et al.*, 2012 High genetic diversity in a remote island population system: *sans* sex. New Phytologist **193**: 1088–1097.

KARLIN, S., J. L. MCGREGOR, and W. BODMER, 1967 The rate of production of recombinants between linked genes in finite populations. Proc. Fifth Berkeley Symp. on Math. Stat. Prob. **4**: 403–414.

KARLIN, S., and H. M. TAYLOR, 1981 *A second course in stochastic processes*. Academic Press, New York.

KEIGHTLEY, P. D., and A. EYRE-WALKER, 2000 Deleterious mutations and the evolution of sex. Science **290**: 331–333.

KEIGHTLEY, P. D., and W. G. HILL, 1987 Directional selection and variation in finite populations. Genetics **117**: 573–582.

KEIGHTLEY, P. D., and M. LYNCH, 2003 Towards a realistic model of mutations affecting fitness. Evolution **57**: 683–685.

KEIGHTLEY, P. D., and S. P. OTTO, 2006 Interference among deleterious mutations favours sex and recombination in finite populations. Nature **443**: 89–92.

KEIGHTLEY, P. D., U. TRIVEDI, M. THOMSON, F. OLIVER, S. KUMAR, *et al.*, 2009 Analysis of the genome sequences of three *Drosophila melanogaster* spontaneous mutation accumulation lines. Genome Res. **19**: 1195–1201.

KELLY, R. P., and S. R. PALUMBI, 2010 Genetic structure among 50 species of the northeastern pacific rocky intertidal community. PLoS ONE **5**: e8594.

KIM, Y., and T. MARUKI, 2011 Hitchhiking effect of a beneficial mutation spreading in a subdivided population. Genetics **189**: 213–226.

KIMURA, M., 1965 Attainment of quasi linkage equilibrium when gene frequencies are changing by natural selection. Genetics **52**: 875–890.

KIMURA, M., 1970 Stochastic processes in population genetics. In K.-I. Kojima, editor, *Mathematical Topics in Population Genetics*, volume 1 of *Biomathematics*. Springer-Verlag; Heidelberg; New York, Berlin, 178–209.

KIMURA, M., 1983 *The neutral theory of molecular evolution*. Cambridge University Press, Cambridge.

KIMURA, M., and T. MARUYAMA, 1966 The mutational load with epistatic gene interactions in fitness. Genetics **54**: 1337–1351.

KIMURA, M., and T. OHTA, 1969 The average number of generations until fixation of a mutant gene in a finite population. Genetics **61**: 763–771.

KIMURA, M., and T. OHTA, 1970 Probability of fixation of a mutant gene in a finite population when selective advantage decreases with time. Genetics **65**: 525–534.

KING, K. C., L. F. DELPH, J. JOKELA, and C. M. LIVELY, 2009 The geographic mosaic of sex and the Red Queen. Curr. Biol. **19**: 1438–1441.

KING, K. C., J. JOKELA, and C. M. LIVELY, 2011a Parasites, sex, and clonal diversity in natural snail populations. Evolution **65**: 1474–1481.

KING, K. C., J. JOKELA, and C. M. LIVELY, 2011b Trematode parasites infect or die in snail hosts. Biol. Lett. **7**: 265–268.

KIRKPATRICK, M., and C. D. JENKINS, 1989 Genetic segregation and the maintenance of sexual reproduction. Nature **339**: 300–301.

KIRKPATRICK, M., T. JOHNSON, and N. BARTON, 2002 General Models of Multilocus Evolution. Genetics **161**: 1727–1750.

KLECKNER, N., 1996 Meiosis: how could it work? Proc. Natl. Acad. Sci. USA **93**: 8167–8174.

KONDRASHOV, A. S., 1982 Selection against harmful mutations in large sexual and asexual populations. Genet. Res. **40**: 325–332.

KONDRASHOV, A. S., 1993 Classification of hypotheses on the advantage of amphimixis. J. Hered. **84**: 372–387.

KONDRASHOV, A. S., and D. HOULE, 1994 Genotype-environment interactions and the estimation of the genomic mutation rate in *Drosophila melanogaster*. Proc. R. Soc. Lond. B. Biol. Sci. **258**: 221–227.

KOSKELLA, B., and C. M. LIVELY, 2009 Evidence for negative frequency-dependent selection during experimental coevolution of a freshwater snail and a sterilizing trematode. Evolution **63**: 2213–2221.

KOUYOS, R. D., O. K. SILANDER, and S. BONHOEFFER, 2007 Epistasis between deleterious mutations and the evolution of recombination. Trends Ecol. Evol. **22**: 308–315.

KUMPULAINEN, T., A. GRAPPUTO, and J. MAPPES, 2004 Parasites and sexual reproduction in psychid moths. Evolution **58**: 1511–1520.

LADLE, R. J., R. A. JOHNSTONE, and O. P. JUDSON, 1993 Coevolutionary dynamics of sex in a metapopulation: escaping the Red Queen. Proc. R. Soc. B **253**: 155–160.

LANDE, R., 1979 Effective deme sizes during long-term evolution estimated from rates of chromosomal rearrangement. Evolution **33**: 234–251.

LEHTONEN, J., M. D. JENNIONS, and H. KOKKO, 2012 The many costs of sex. Trends Ecol. Evol. **27**: 172–178.

LENORMAND, T., 2002 Gene flow and the limits to natural selection. Trends Ecol. Evol. **17**: 183–189.

LENORMAND, T., and S. P. OTTO, 2000 The evolution of recombination in a heterogeneous environment. Genetics **156**: 423–438.

LEWONTIN, R. C., and J. KRAKAUER, 1973 Distribution of gene frequency as a test of the theory of the selective neutrality of polymorphisms. Genetics **74**: 175–195.

LIVELY, C. M., 1987 Evidence from a New Zealand snail for the maintenance of sex by parasitism. Nature **328**: 519–521.

LIVELY, C. M., 2009 The maintenance of sex: host-parasite coevolution with density-dependent virulence. J. Evol. Biol. **22**: 2086–2093.

LIVELY, C. M., 2010a An epidemiological model of host-parasite coevolution and sex. J. Evol. Biol. **23**: 1490–1497.

LIVELY, C. M., 2010b Parasite virulence, host life history, and the costs and benefits of sex. Ecology **91**: 3–6.

LIVELY, C. M., and D. G. LLOYD, 1990 The cost of biparental sex under individual selection. Am. Nat. **135**: 489–500.

LOCKHART, A. B., P. H. THRALL, and J. ANTONOVICS, 1996 Sexually transmitted diseases in animals: ecological and evolutionary implications. Biol. Rev. Camb. Philos. Soc. **71**: 415–471.

LOEWE, L., and B. CHARLESWORTH, 2006 Inferring the distribution of mutational effects on fitness in *Drosophila*. Biol. Lett. **2**: 426–430.

LYNCH, M., R. BURGER, D. BUTCHER, and W. GABRIEL, 1993 The mutational meltdown in asexual populations. J. Hered. **84**: 339–344.

LYNCH, M., and W. G. HILL, 1986 Phenotypic evolution by neutral mutation. Evolution **40**: 915–935.

LYNCH, M., and B. WALSH, 1998 *Genetics and analysis of quantitative traits*. Sinauer, Sunderland, Mass.

LYTHGOE, K. A., 2000 The coevolution of parasites with host-acquired immunity and the evolution of sex. Evolution **54**: 1142–1156.

MALMBERG, R. L., 1977 The evolution of epistasis and the advantage of recombination in populations of bacteriophage T4. Genetics **86**: 607–621.

MARK WELCH, D. B., J. L. MARK WELCH, and M. MESELSON, 2008 Evidence for degenerate tetraploidy in bdelloid rotifers. Proc. Natl. Acad. Sci. USA **105**: 5145–5149.

MARK WELCH, D. B., and M. MESELSON, 2000 Evidence for the evolution of bdelloid rotifers without sexual reproduction or genetic exchange. Science **288**: 1211–1215.

MARTENS, E. A., and O. HALLATSCHEK, 2011 Interfering waves of adaptation promote spatial mixing. Genetics **189**: 1045–1060.

MARTIN, G., S. P. OTTO, and T. LENORMAND, 2006 Selection for recombination in structured populations. Genetics **172**: 593–609.

MARUYAMA, T., 1970 On the fixation probability of mutant genes in a subdivided population. Genet. Res. **15**: 221–225.

MAY, R. M., and R. M. ANDERSON, 1983 Epidemiology and genetics in the coevolution of parasites and hosts. Proc. R. Soc. Lond. B. Biol. Sci. **219**: 281–313.

MAYNARD SMITH, J., 1971 What use is sex? J. Theor. Biol. **30**: 319–335.

MAYNARD SMITH, J., 1978 *The evolution of sex*. Cambridge University Press, Cambridge; New York.

MAYNARD SMITH, J., 1988a The evolution of recombination. In R. E. Michod and B. R. Levin, editors, *The Evolution of Sex*. Sinauer Press, Sunderland, Massachusetts, 106–125.

MAYNARD SMITH, J., 1988b Selection for recombination in a polygenic model - the mechanism. Genet. Res. **51**: 59–63.

MAYNARD SMITH, J., and J. HAIGH, 1974 The hitch-hiking effect of a favourable gene. Genet. Res. **23**: 23–35.

MCVEAN, G. A. T., and B. CHARLESWORTH, 2000 The effects of Hill-Robertson interference between weakly selected mutations on patterns of molecular evolution and variation. Genetics **155**: 929–944.

MCVEAN, G. A. T., S. R. MYERS, S. HUNT, P. DELOUKAS, D. R. BENTLEY, *et al.*, 2004 The fine-scale structure of recombination rate variation in the human genome. Science **304**: 581–584.

MICHAELS, H. J., and F. A. BAZZAZ, 1986 Resource allocation and demography of sexual and apomictic *Antennaria parlinii*. Ecology **67**: 27–36.

MILLER, M. P., D. W. BLINN, and P. KEIM, 2002 Correlations between observed dispersal capabilities and patterns of genetic differentiation in populations of four aquatic insect species from the Arizona White Mountains, U.S.A. Freshw. Biol. **47**: 1660–1673.

MORJAN, C. L., and L. H. RIESEBERG, 2004 How species evolve collectively: implications of gene flow and selection for the spread of advantageous alleles. Mol. Ecol. **13**: 1341–1356.

MORRAN, L. T., M. D. PARMENTER, and P. C. PHILLIPS, 2009 Mutation load and rapid adaptation favour outcrossing over self-fertilization. Nature **462**: 350–352.

MORRAN, L. T., O. G. SCHMIDT, I. A. GELARDEN, R. C. PARRISH, and C. M. LIVELY, 2011 Running with the Red Queen: Host-parasite coevolution selects for biparental sex. Science **333**: 216–218.

MOSTOWY, R., M. SALATHÉ, R. D. KOUYOS, and S. BONHOEFFER, 2010 On the evolution of sexual reproduction in hosts coevolving with multiple parasites. Evolution **64**: 1644–1656.

MULLER, H. J., 1932 Some genetic aspects of sex. Am. Nat. **66**: 118–138.

MULLER, H. J., 1964 The relation of recombination to mutational advance. Mutat. Res. **1**: 2–9.

NEI, M., 1967 Modification of linkage intensity by natural selection. Genetics **57**: 625–641.

NEI, M., 1973 Analysis of gene diversity in subdivided populations. Proc. Natl. Acad. Sci. USA **70**: 3321–3323.

NEIMAN, M., G. HEHMAN, J. T. MILLER, J. LOGSDON, JOHN M., and D. R. TAYLOR, 2010 Accelerated mutation accumulation in asexual lineages of a freshwater snail. Mol. Biol. Evol. **27**: 954–963.

NEIMAN, M., J. JOKELA, C. M. LIVELY, and L. KATZ, 2005 Variation in asexual lineage age in *Potamopyrgus antipodarum*, a New Zealand snail. Evolution **59**: 1945–1952.

NIELSEN, R., 2005 Molecular signals of natural selection. Annu. Rev. Genet. **39**: 197–218.

OBBARD, D. J., J. J. WELCH, K.-W. KIM, and F. M. JIGGINS, 2009 Quantifying adaptive evolution in the *Drosophila* immune system. PLoS Genet. **5**: e1000698.

OHTA, T., and M. KIMURA, 1975 The effect of selected linked locus on heterozygosity of neutral alleles (the hitch-hiking effect). Genet. Res. **25**: 313–325.

OLIVEIRA, V. M. D., J. K. D. SILVA, and P. R. A. CAMPOS, 2008 Epistasis and the selective advantage of sex and recombination. Phys. Rev. E: Stat., Nonlinear, Soft Matter Phys. **78**: 031905.

ORR, H. A., 2010 The population genetics of beneficial mutations. Philos. Trans. R. Soc. Lond. B Biol. Sci. **365**: 1195–1201.

OTTO, S. P., 2003 The advantages of segregation and the evolution of sex. Genetics **164**: 1099–1118.

OTTO, S. P., 2009 The evolutionary enigma of sex. Am. Nat. **174**: S1–S14.

OTTO, S. P., and N. H. BARTON, 1997 The evolution of recombination: Removing the limits to natural selection. Genetics **147**: 879–906.

OTTO, S. P., and N. H. BARTON, 2001 Selection for recombination in small populations. Evolution **55**: 1921–1931.

OTTO, S. P., and M. W. FELDMAN, 1997 Deleterious mutations, variable epistatic interactions, and the evolution of recombination. Theor. Popul. Biol. **51**: 134–147.

OTTO, S. P., and T. LENORMAND, 2002 Resolving the paradox of sex and recombination. Nat. Rev. Genet. **3**: 252–261.

OTTO, S. P., and S. L. NUISMER, 2004 Species interactions and the evolution of sex. Science **304**: 1018–1020.

PAETKAU, D., S. C. AMSTRUP, E. W. BORN, W. CALVERT, A. E. DEROCHER, *et al.*, 1999 Genetic structure of the world's polar bear populations. Mol. Ecol. **8**: 1571–1584.

PALAND, S., and M. LYNCH, 2006 Transitions to asexuality result in excess amino acid substitutions. Science **311**: 990–992.

PANNELL, J. R., M. E. DORKEN, and S. M. EPPLEY, 2005 'Haldane's Sieve' in a metapopulation: sifting through plant reproductive polymorphisms. Trends Ecol. Evol. **20**: 374–379.

PATWA, Z., and L. M. WAHL, 2008 The fixation probability of beneficial mutations. J. R. Soc. Interface **5**: 1279–1289.

PECK, J. R., 1993 Frequency-dependent selection, beneficial mutations, and the evolution of sex. Proc. R. Soc. B **254**: 87–92.

PECK, J. R., 1994 A ruby in the rubbish: Beneficial mutations, deleterious mutations and the evolution of sex. Genetics **137**: 597–606.

PECK, J. R., G. BARREAU, and S. C. HEATH, 1997 Imperfect genes, Fisherian mutation and the evolution of sex. Genetics **145**: 1171–1199.

PECK, J. R., and D. WAXMAN, 2000 What's wrong with a little sex? J. Evol. Biol. **13**: 63–69.

PECK, J. R., J. YEARSLEY, and G. BARREAU, 1999 The maintenance of sexual reproduction in a structured population. Proc. R. Soc. B **266**: 1857–1863.

PÉREZ-ESPONA, S., F. PÉREZ-BARBERÍA, J. MCLEOD, C. JIGGINS, I. GORDON, *et al.*, 2008 Landscape features affect gene flow of Scottish Highland red deer (*Cervus elaphus*). Mol. Ecol. **17**: 981–996.

PETERS, A. D., and C. M. LIVELY, 1999 The Red Queen and fluctuating epistasis: A population genetic analysis of antagonistic coevolution. Am. Nat. **154**: 393–405.

PETERS, A. D., and C. M. LIVELY, 2007 Short- and long-term benefits and detriments to recombination under antagonistic coevolution. J. Evol. Biol. **20**: 1206–1217.

PIÁLEK, J., and N. H. BARTON, 1997 The spread of an advantageous allele across a barrier: The effects of random drift and selection against heterozygotes. Genetics **145**: 493–504.

POOL, J. E., and C. F. AQUADRO, 2006 History and structure of Sub-Saharan populations of *Drosophila melanogaster*. Genetics **174**: 915–929.

POON, A., and L. CHAO, 2004 Drift increases the advantage of sex in RNA bacteriophage Φ6. Genetics **166**: 19–24.

POON, A., and S. P. OTTO, 2000 Compensating for our load of mutations: Freezing the mutational meltdown. Evolution **54**: 1467–1479.

PRESGRAVES, D. C., 2005 Recombination enhances protein adaptation in *Drosophila melanogaster*. Curr. Biol. **15**: 1651–1656.

PRICE, G. R., 1972 Fisher's 'fundamental theorem' made clear. Ann. Hum. Genet. **36**: 129–140.

RALPH, P., and G. COOP, 2010 Parallel adaptation: One or many waves of advance of an advantageous allele? Genetics **186**: 647–668.

REDFIELD, R. J., 1988 Evolution of bacterial transformation: Is sex with dead cells ever better than no sex at all? Genetics **119**: 213–221.

REDFIELD, R. J., 1993 Evolution of Natural Transformation: Testing the DNA Repair Hypothesis in *Bacillus subtilis* and *Haemophilus influenzae*. Genetics **133**: 755–761.

RICE, W. R., 1999 Free content genetic polarization: unifying theories for the adaptive significance of recombination. J. Evol. Biol. **12**: 1047–1049.

RICE, W. R., and A. K. CHIPPINDALE, 2001 Sexual recombination and the power of natural selection. Science **294**: 555–559.

RIGINOS, C., K. E. DOUGLAS, Y. JIN, D. F. SHANAHAN, and E. A. TREML, 2011 Effects of geography and life history traits on genetic differentiation in benthic marine fishes. Ecography .

ROBERTSON, A., 1961 Inbreeding in artificial selection programmes. Genet. Res. **2**: 189–194.

ROEDER, A. D., R. K. MARSHALL, A. J. MITCHELSON, T. VISAGATHILAGAR, P. A. RITCHIE, *et al.*, 2001 Gene flow on the ice: genetic differentiation among Adélie penguin colonies around Antarctica. Mol. Ecol. **10**: 1645–1656.

ROZE, D., 2009 Diploidy, population structure, and the evolution of recombination. Am. Nat. **174**: S79–S94.

ROZE, D., and N. H. BARTON, 2006 The Hill-Robertson effect and the evolution of recombination. Genetics **173**: 1793–1811.

ROZE, D., and R. E. MICHOD, 2010 Deleterious mutations and selection for sex in finite diploid populations. Genetics **184**: 1095–1112.

ROZE, D., and F. ROUSSET, 2003 Selection and drift in subdivided populations: A straightforward method for deriving diffusion approximations and applications involving dominance, selfing and local extinctions. Genetics **165**: 2153–2166.

SALATHÉ, M., R. D. KOUYOS, and S. BONHOEFFER, 2008a The state of affairs in the kingdom of the Red Queen. Trends Ecol. Evol. **23**: 439–445.

SALATHÉ, M., R. D. KOUYOS, and S. BONHOEFFER, 2009 On the causes of selection for recombination underlying the Red Queen hypothesis. Am. Nat. **174**: S31–S42.

SALATHÉ, M., R. D. KOUYOS, R. R. REGOES, and S. BONHOEFFER, 2008b Rapid parasite adaptation drives selection for high recombination rates. Evolution **62**: 295–300.

SALATHÉ, M., R. SALATHÉ, P. SCHMID-HEMPEL, and S. BONHOEFFER, 2006 Mutation accumulation in space and the maintenance of sexual reproduction. Ecol. Lett. **9**: 941–946.

SANTIAGO, E., and A. CABALLERO, 2005 Variation after a selective sweep in a subdivided population. Genetics **169**: 475–483.

SATTATH, S., E. ELYASHIV, O. KOLODNY, Y. RINOTT, and G. SELLA, 2011 Pervasive adaptive protein evolution apparent in diversity patterns around amino acid substitutions in *Drosophila simulans*. PLoS Genet. **7**: e1001302.

SCHNEIDER, A., B. CHARLESWORTH, A. EYRE-WALKER, and P. D. KEIGHTLEY, 2011 A method for inferring the rate of occurrence and fitness effects of advantageous mutations. Genetics **189**: 1427–1437.

SCHOUSTRA, S., H. D. RUNDLE, R. DALI, and R. KASSEN, 2010 Fitness-associated sexual reproduction in a filamentous fungus. Curr. Biol. **20**: 1350–1355.

SELLA, G., D. A. PETROV, M. PRZEWORSKI, and P. ANDOLFATTO, 2009 Pervasive natural selection in the *Drosophila* genome? PLoS Genet. **5**: e1000495.

SHAPIRO, J. A., W. HUANG, C. ZHANG, M. J. HUBISZ, J. LU, *et al.*, 2007 Adaptive genic evolution in the *Drosophila* genomes. Proc. Natl. Acad. Sci. USA **104**: 2271–2276.

SHAW, F. H., C. J. GEYER, and R. G. SHAW, 2002 A comprehensive model of mutations affecting fitness and interferences for *Arabidopsis thaliana*. Evolution **56**: 453–463.

SHAW, P., G. PIERCE, and P. BOYLE, 1999 Subtle population structuring within a highly vagile marine invertebrate, the veined squid *Loligo forbesi*, demonstrated with microsatellite DNA markers. Mol. Ecol. **8**: 407–417.

SHIGESADA, N., and K. KAWASAKI, 1997 *Biological Invasions: Theory and Practice*. Oxford series in ecology and evolution. Oxford University Press, Oxford.

SILLER, S., 2001 Sexual selection and the maintenance of sex. Nature **411**: 689–692.

SIMMONS, M. J., and J. F. CROW, 1977 Mutations affecting fitness in *Drosophila* populations. Annu. Rev. Genet. **11**: 49–78.

SIMON, J.-C., F. DELMOTTE, C. RISPE, and T. CREASE, 2003 Phylogenetic relationships between parthenogens and their sexual relatives: the possible routes to parthenogenesis in animals. Biol. J. Linn. Soc. **79**: 151–163.

SINGH, R. S., and L. R. RHOMBERG, 1987 A Comprehensive Study of Genic Variation in Natural Populations of *Drosophila melanogaster*. II. Estimates of Heterozygosity and Patterns of Geographic Differentiation. Genetics **117**: 255–271.

SLATKIN, M., 1976 The rate of spread of an advantageous allele in a subdivided population. In S. Karlin and E. Nevo, editors, *Population Genetics and Ecology*. Academic Press, New York, 767–780.

SLATKIN, M., 1981 Fixation probabilities and fixation times in a subdivided population. Evolution **35**: 477–488.

SLATKIN, M., 1995 A measure of population subdivision based on microsatellite allele frequencies. Genetics **139**: 457–462.

SLATKIN, M., and T. H. E. WIEHE, 1998 Genetic hitch-hiking in a subdivided population. Genet. Res. **71**: 155–160.

SMITH, N. G. C., and A. EYRE-WALKER, 2002 Adaptive protein evolution in *Drosophila*. Nature **415**: 1022–1024.

STENBERG, P., and A. SAURA, 2009 Cytology of asexual animals. In I. Schön, K. Martens and P. Dijk, editors, *Lost Sex: The Evolutionary Biology of Parthenogenesis*. Springer Netherlands, Dordrecht, 63–74.

STEPHAN, W., T. H. E. WIEHE, and M. W. LENZ, 1992 The effect of strongly selected substitutions on neutral polymorphism: analytical results based on diffusion theory. Theor. Popul. Biol. **41**: 237–254.

THE CHIMPANZEE SEQUENCING AND ANALYSIS CONSORTIUM, 2005 Initial sequence of the chimpanzee genome and comparison with the human genome. Nature **437**: 69–87.

THOMSON, G., 1977 The effect of a selected locus on linked neutral loci. Genetics **85**: 753–788.

VRIJENHOEK, R. C., 1998 Animal clones and diversity. BioScience **48**: 617–628.

VUILLEUMIER, S., J. M. YEARSLEY, and N. PERRIN, 2008 The fixation of locally beneficial alleles in a metapopulation. Genetics **178**: 467–475.

WALLACE, B., 1975 Hard and soft selection revisited. Evolution **29**: 465–473.

WEIR, B. S., and C. C. COCKERHAM, 1984 Estimating F-Statistics for the analysis of population structure. Evolution **38**: 1358–1370.

WEISMANN, A., 1887 On the signification of the polar globules. Nature **36**: 607–609.

WEST, S., C. LIVELY, and A. READ, 1999 A pluralist approach to sex and recombination. J. Evol. Biol. **12**: 1003–1012.

WHITLOCK, M. C., 2002 Selection, load and inbreeding depression in a large metapopulation. Genetics **160**: 1191–1202.

WHITLOCK, M. C., 2003 Fixation probability and time in subdivided populations. Genetics **164**: 767–779.

WHITLOCK, M. C., 2011 G'_{st} and D do not replace F_{st}. Mol. Ecol. **20**: 1083–1091.

WHITLOCK, M. C., and R. GOMULKIEWICZ, 2005 Probability of fixation in a heterogeneous environment. Genetics **171**: 1407–1417.

WHITLOCK, M. C., and D. E. MCCAULEY, 1999 Indirect measures of gene flow and migration: $F_{st} \neq 1/(4Nm + 1)$. Heredity **82**: 117–125.

WILLIAMS, G. C., 1975 *Sex and evolution*. Princeton University Press, Princeton.

WILLIAMSON, S. H., M. J. HUBISZ, A. G. CLARK, B. A. PAYSEUR, C. D. BUSTAMANTE, *et al.*, 2007 Localizing recent adaptive evolution in the human genome. PLoS Genet. **3**: e90.

WOLFRAM RESEARCH, I., 2010 *Mathematica Edition: Version 8.0*. Wolfram Research, Inc., Champaign, Illinois.

WORLEY, K., C. STROBECK, S. ARTHUR, J. CAREY, H. SCHWANTJE, *et al.*, 2004 Population genetic structure of North American thinhorn sheep (*Ovis dalli*). Mol. Ecol. **13**: 2545–2556.

WRIGHT, S., 1931 Evolution in Mendelian populations. Genetics **16**: 97–159.

WRIGHT, S., 1951 The genetical structure of populations. Ann. Eugen. **15**: 323–354.

WRIGHT, S. I., and P. ANDOLFATTO, 2008 The impact of natural selection on the genome: Emerging patterns in *Drosophila* and *Arabidopsis*. Annu. Rev. Ecol. Systematics **39**: 193–213.

YU, F., and A. M. ETHERIDGE, 2008 Rate of adaptation of large populations. In P. Pontarotti, editor, *Evolutionary Biology from Concept to Application*. Springer-Verlag, Berlin, 3–27.

YU, F., and A. M. ETHERIDGE, 2010 The fixation probability of two competing beneficial mutations. Theor. Popul. Biol. **78**: 36–45.

YU, F., A. M. ETHERIDGE, and C. CUTHBERTSON, 2010 Asymptotic behavior of the rate of adaptation. Ann. Appl. Probab. **20**: 978–1004.

YUKILEVICH, R., T. L. TURNER, F. AOKI, S. V. NUZHDIN, and J. R. TRUE, 2010 Patterns and Processes of Genome-Wide Divergence Between North American and African *Drosophila melanogaster*. Genetics **186**: 219–239.

ZEYL, C., and G. BELL, 1997 The advantage of sex in evolving yeast populations. Nature **388**: 465–468.

ZHANG, L., and W.-H. LI, 2005 Human SNPs reveal no evidence of frequent positive selection. Mol. Biol. Evol. **22**: 2504–2507.

Appendix

DOI: 10.1534/genetics.109.112920

The Role of Advantageous Mutations in Enhancing the Evolution of a Recombination Modifier

Matthew Hartfield,*[,1] Sarah P. Otto[†] and Peter D. Keightley*

**Institute of Evolutionary Biology, School of Biological Sciences, University of Edinburgh, Edinburgh EH9 3JT, United Kingdom and [†]Department of Zoology, University of British Columbia, Vancouver, British Columbia V6T 1Z4, Canada*

Manuscript received December 10, 2009
Accepted for publication February 3, 2010

ABSTRACT

Although the evolution of recombination is still a major problem in evolutionary genetics, recent theoretical studies have shown that recombination can evolve by breaking down interference ("Hill–Robertson effects") among multiple loci. This leads to selection on a recombination modifier in a population subject to recurrent deleterious mutation. Here, we use computer simulations to investigate the evolution of a recombination modifier under three different scenarios of recurrent mutation in a finite population: (1) mutations are deleterious only, (2) mutations are advantageous only, and (3) there is a mixture of deleterious and advantageous mutations. We also investigate how linkage disequilibrium, the strength of selection acting on a modifier, and effective population size change under the different scenarios. We observe that adding even a small number of advantageous mutations increases the fixation rate of modifiers that increase recombination, especially if the effects of deleterious mutations are weak. However, the strength of selection on a modifier is less than the summed strengths had there been deleterious mutations only and advantageous mutations only.

SEX and recombination between genomes are ubiquitous in nature, yet explaining their evolution has not proved to be easy (see recent reviews by Hadany and Comeron 2008 and Otto 2009). Recombination leads to the breakup of beneficial gene combinations (Barton and Charlesworth 1998), implying that offspring may suffer a recombination load (Charlesworth and Charlesworth 1975). Extra costs are incurred if sexual reproduction is also considered, such as the famous "twofold cost" (Maynard Smith 1978); sexual offspring need two parents whereas asexuals need one, so the latter can outgrow and outcompete sexuals.

Considering all of the associated costs, it has proved difficult to explain why recombination and sex are so common among eukaryotes. One hypothesis is that recombination breaks down "Hill–Robertson effects" in asexuals (Hill and Robertson 1966), which otherwise impede the response to selection. Hill–Robertson effects are the manifestation of many phenomena (discussed further in Charlesworth *et al.* 2009), including hitchhiking (Maynard Smith and Haigh 1974), background selection (Charlesworth *et al.* 1993), and the accumulation of deleterious mutations by Muller's ratchet (Felsenstein 1974; Muller 1964). Interference generates negative linkage disequilibrium (*i.e.*, the accumulation of good alleles on bad genetic backgrounds), which reduces genetic variation in fitness compared to a population without this linkage disequilibrium (Barton 2009). Interference also reduces the effective population size, N_e (Robertson 1961; Comeron *et al.* 2008), because offspring from the same, fittest, lineages tend to be favored. Recombination can increase the genetic variance in overall fitness, which can improve the response to selection (Fisher 1930; Maynard Smith 1988). If a modifier for increased recombination facilitates the production of fitter offspring in this way, then it has an indirect selective advantage and increases in frequency by virtue of being associated with fitter genotypes, in line with Weismann's classic theory on the evolution of sex and recombination (Weismann 1887; Burt 2000).

Research into the Hill–Robertson effect has increased in recent years, with the development of analytical frameworks to study effects of drift at multiple loci. By extending earlier models that focused on selection alone (Barton 1995; Otto and Feldman 1997), recent work has assessed how linkage disequilibrium, created by genetic drift and interference with selection, drives the evolution of a recombination modifier (Barton and Otto 2005). Indirect selection on a modifier also arises because recombination increases the probability that beneficial mutations establish within a population, and the strength of this selection has been modeled

Supporting information is available online at http://www.genetics.org/cgi/content/full/genetics.109.112920/DC1.

[1]*Corresponding author:* Ashworth Laboratories, King's Bldgs., University of Edinburgh, W. Mains Rd., Edinburgh EH9 3JT, United Kingdom. E-mail: m.hartfield@sms.ed.ac.uk

using multitype branching processes (Otto and Barton 1997; Roze and Barton 2006).

Simultaneously, the ability to simulate large numbers of linked loci has increased, making it possible to evaluate the importance of the Hill–Robertson effect with selection acting across a genome. Simulations have demonstrated how breaking down interference can offer substantial selection on a modifier of recombination. Otto and Barton (2001), for example, showed that if a recombination modifier acts on loci experiencing directional selection, the effects of drift (which creates interference between loci) account for more selection on recombination than the effects of epistasis, in both 3-locus and 11-locus simulations. With only 3 loci, however, a modifier was not favored in populations of $N > 10{,}000$ chromosomes without epistasis. Subsequently, Iles *et al.* (2003) demonstrated that as the number of loci under directional selection increases, strong selection for recombination occurs in even larger populations, including the largest population size considered (100,000 haploid individuals). Furthermore, with population structure, breaking down Hill–Robertson effects remains important for modifier evolution even in infinitely large populations, with finite deme size (Martin *et al.* 2006). Similarly, Keightley and Otto (2006) showed that a recombination modifier is strongly favored in a population subject to recurrent deleterious mutation. The effects of drift again overwhelmed epistasis, and the dramatic reduction in the effective population size at a neutral site, N_e, in asexual populations highlighted how Hill–Robertson effects impede natural selection (*e.g.*, reductions from $N = 50{,}000$ to $N_e \approx 100$, if selection acting against deleterious mutants equaled 0.01 and there was complete linkage between loci). Selection on a sex modifier was also sufficiently strong that it could overcome a twofold cost, but only if sex was initially rare and the modifier led to modest increases in the frequency of sex.

These advantages of recombination have been supported by experiments demonstrating that recombining populations are more responsive to selection. A study by Malmberg (1977) found that allowing the bacteriophage T4 to exchange segments of its genome improved its rate of adaptation. More recent studies provide evidence that recombination increases the realized selection strength and fixation rate of new mutants in *Drosophila melanogaster* (Rice and Chippindale 2001); regions of the genome lacking in recombination in Drosophila show signs that selection has been impeded, whereas regions that have normal levels of recombination appear to adapt more quickly (Presgraves 2005; Betancourt *et al.* 2009; Charlesworth *et al.* 2009); sex overcomes clonal interference in *Chlamydomonas reinhardtii*, which then accelerates adaptation (Colegrave 2002); sex in stressed environments of yeast increases population variance in fitness and the response to selection (Goddard *et al.* 2005); genetic drift induced by population bottlenecks in the RNA bacteriophage Φ6 hampers the response to selection to a greater extent in asexuals then sexuals (Poon and Chao 2004); recombining populations of *Escherichia coli* are better able to break down interference between a known beneficial allele and other sites under selection, thereby increasing the rate of fixation of the fitter allele (Cooper 2007); and *Caenorhabditis elegans* evolves outcrossing if a population is subject to an increased mutation rate or the presence of a pathogen, indicating that sex improves the response to adaptation (Morran *et al.* 2009).

Recently, a "pluralist" framework was proposed (West *et al.* 1999), arguing that multiple mechanisms work together to facilitate the widespread evolution of genetic recombination. A recent example investigating this was undertaken by Oliveira *et al.* (2008), which aimed to gauge what degree of selection strengths acting on deleterious mutants drove the evolution of recombination via the process of background selection or through Muller's ratchet.

Here, we use computer simulations to extend the work of Keightley and Otto (2006) by considering both deleterious and advantageous mutations arising throughout the genome, in the spirit of a more pluralist approach to the Hill–Robertson effect. Whereas Keightley and Otto (2006) demonstrated that genetic recombination is selectively favored if multiple linked loci within the genome are subject to recurrent deleterious mutation, Iles *et al.* (2003) demonstrated that similar advantages to a modifier occur if multiple linked loci are subject to recurrent advantageous mutations. This motivates the question: In genomes subject to both deleterious and advantageous mutations, is recombination favored more compared to a population subject to only one type of mutation and, if so, by how much? Specifically, we investigate whether the benefits of a recombination modifier in the presence of deleterious and advantageous mutations are additive. That is, if the modifier has selection advantage s_{Md} if just recurrent deleterious mutations occur at rate U_d, and advantage s_{Ma} if just advantageous mutations occur at rate U_a, then with both types of mutation occurring at a total rate of $U_d + U_a$, additivity implies a selective advantage of the modifier equal to $s_{Md} + s_{Ma}$. This is a reasonable null hypothesis to start with, and it allows us to test for the presence of interference between different types of mutation.

At sites under selection, the extent to which nucleotide substitutions are driven by positive selection or occur despite negative selection has been the topic of long debate. The proportion of advantageous mutations is likely to depend strongly on the match between the species and its current environment (for a recent review see Eyre-Walker 2006). Comparing the chimpanzee and human genomes suggests that hominids have experienced little adaptive evolution at the molecular level (Chimpanzee Sequencing and Analysis Consortium

2005; Zhang and Li 2005) [although low rates of adaptive molecular evolution inferred in these species could be a consequence of population bottlenecks, which would downwardly bias estimates (Eyre-Walker and Keightley 2009)]. On the other hand, Bierne and Eyre-Walker (2004) inferred that ~45% of amino acid substitutions in Drosophila are a consequence of adaptive evolution. This equates to one substitution in the genome every 45 years (~450 generations) (Smith and Eyre-Walker 2002), although these estimates are subject to discussion (Sella *et al.* 2009). Here, we use rates of positive mutation based on these data to determine the role that adaptive mutations might play in the evolution of genetic recombination, especially when there are background deleterious mutations as well.

METHODS

Simulation of a recombination modifier: The simulations start with a population of N mutant-free haploid chromosomes, each consisting of 100 equally spaced linked loci subject to recurrent mutation. A new generation is created by selection, recombination (if present), and mutation to produce N offspring.

Three scenarios are investigated: mutants are exclusively deleterious (as in Keightley and Otto 2006); mutants are exclusively advantageous (similar to Iles *et al.* 2003, although they considered standing variation only); or a proportion $x = k/s_a$ of mutants are advantageous and $1 - x$ are deleterious, for k a numerical constant and s_a the selection acting on an advantageous mutant. The function $x = k/s_a$ reflects the assumption that strongly advantageous mutants are less likely to appear than weakly selected ones (Andolfatto 2007; Jensen *et al.* 2008). The number of mutants is chosen from a Poisson distribution with mean U, except for the case where all spontaneous mutations are advantageous. Specifically, we assume that advantageous mutations are always a small proportion, x, of all mutations that occur. Thus, when only advantageous mutations are present, they occur at a rate Ux (this is equivalent to setting $s_d = 0$ in the case with both advantageous and deleterious mutants). If both advantageous and deleterious mutations are present, the overall mutation rate is not quite the sum of mutation rates from the separate scenarios; the overall mutation rate is U whereas the summed rate if mutants are deleterious only and advantageous only equals $U + Ux = (1 + x)U$. However, x is assumed to be small (it is always <3%) and simulations with a total deleterious and advantageous mutation rate of $(1 + x)U$ give indistinguishable estimates of s_M, the selection strength on the modifier, compared to simulations with overall mutation rate U (results not shown).

In all scenarios, each site is equally likely to acquire a new mutation. Fitness effects of loci are multiplicative, with advantageous mutants having fixed fitness effects s_a and deleterious mutants have fitness effects of s_d. Thus with y advantageous mutants and z deleterious mutants, the fitness of a haploid individual equals $(1 + s_a)^y (1 - s_d)^z$. Fixed fitness effects are used to speed up simulations. Epistasis between mutations on the log-fitness scale is assumed to be absent, so that any increase in fixation rate of modifier mutations can be attributed to Hill–Robertson effects. Every 500 generations, the number of mutants is normalized; the number present in a single chromosome is reduced by the minimum number of advantageous or deleterious mutants that any haploid individual possesses so that the smallest number present in a single chromosome equals zero.

Except where noted, the population initially lacks genetic crossing over. To produce the next generation, a parent is chosen with replacement, with probability proportional to its fitness. This is then cloned and a number of mutants sampled from a Poisson distribution are added to produce an offspring. This is repeated N times until the population is replenished. A burn-in of $5N$ such generations is run to allow the population variance to approach a steady state. The state of the burn-in population is then saved, and a recombination modifier is introduced at a randomly selected position on a randomly selected chromosome. The processes of selection, recombination, and mutation are then repeated, except that two haploid parents are mated to allow crossovers to occur. The modifier increases the (Poisson) mean number of crossover events per chromosome during reproduction from $L = 0$ to $L = 0.1$ if it is present as a homozygote (and half that if it is heterozygous). The new modifier allele is then tracked until it is fixed or lost from the population. The process of introducing a single modifier mutation is repeated $5N$ times for each saved burn-in population and the total number of fixations is divided by $5N$ to obtain the fixation probability u. The statistic used to determine the selective advantage of the modifier is u/u^*, where $u^* = 1/N$, the fixation probability of a neutral mutation (Kimura 1983). The above constituted one "run" to produce a single statistic. Each run is executed 100 times from separate burn-ins to produce a distribution of fixation probabilities.

Parameter values used: The per-chromosome mutation rate (if mutants are solely deleterious or deleterious and advantageous) is set to either $U = 0.1$ or $U = 0.5$, which is in the range of estimated deleterious mutation rates per chromosome in Drosophila (Halligan and Keightley 2006; Haag-Liautard *et al.* 2007; Keightley *et al.* 2009). These values should also be similar to the joint deleterious and advantageous mutation rates, since selected mutants are believed to be mainly deleterious (Crow 1970). x, the proportion of mutants that are advantageous, is set to k/s_a with $k = 0.00023$. This value of k is chosen so that there was, on average,

one substitution every 450 generations (a rate inferred for the Drosophila genome by Bierne and Eyre-Walker 2004), in simulations that we conducted with a small population ($N = 100$), a low mutation rate ($U = 0.1$), with medium-strength advantageous and deleterious mutations both present ($s_a = s_d = 0.025$), and complete linkage between loci. This value of k is then used in all simulations investigated; however, this will lead to higher rates of substitution occurring in simulations with large population sizes or mutation rates.

Values of s_a are set to 0.01, 0.025, or 0.05. We wanted to ensure that $Ns_a \geq 1$ for all $N \geq 100$, so that the fate of mutations is not determined by the action of drift alone, even if Hill–Robertson effects are absent (Kimura 1983). These values are therefore somewhat higher then those obtained from analysis of amino acid substitution data from Drosophila, although there is some overlap (see reviews by Sella *et al.* 2009 and Wright and Andolfatto 2008).The appearance of strong adaptive mutations is best representative of advantageous mutants occurring at nonsynonymous sites, where the substitution rate and selection strength are highest (Andolfatto 2005).

We investigated a wide range of s_d values, from 0 to 0.05. Again precise values of s_d are hard to obtain from observations; smaller values of s_d investigated match up with estimates obtained from Loewe and Charlesworth (2006); however, García-Dorado *et al.* (1999) found a mean s_d of $\sim$0.2. This high value may have resulted from simplifications used in the Bateman–Mukai inference method (Lynch and Walsh 1988). One should be aware though that due to the strongly leptokurtic distributions of s_d found empirically, there is a great deal of variance around such estimates and many s_d values would be lower than those used in our simulations.

Measuring linkage disequilibrium for an asexual and recombining population: The log fitness associated with a chromosome is additive in these simulations, so standard models of the expression of phenotypic quantitative traits can be used to measure differences in variance (Bulmer 1976, 1980; Keightley and Hill 1987). To measure linkage disequilibrium in an asexual population, the frequency of each individual mutant is tracked. A "garbage collection" routine is executed every 10 generations to clear memory; mutants that have become either fixed or lost are removed from the population, and a note is kept of how many new mutant alleles have fixed. There is a burn-in of $5N$ generations, after which the mean linkage disequilibrium is measured over $5N$ generations: linkage disequilibrium (LD) = $V_A - V_g$ for genetic variance V_A and genic variance V_g of the log fitness (Keightley and Hill 1987). LD is the contribution to the genic variance of log fitness due to multilocus linkage disequilibrium (Bulmer 1980) and can be computed as above using the following terms for V_A and V_g,

$$V_A = \frac{(\sum_{i=1}^{N} w_i^2)}{N} - \frac{(\sum_{i=1}^{N} w_i)^2}{N^2} \quad (1)$$

$$V_g = s_a^2\left(\sum_{j=1}^{m} \tilde{y}_j(1-\tilde{y}_j)\right) + s_d^2\left(\sum_{j=1}^{m} \tilde{z}_j(1-\tilde{z}_j)\right), \quad (2)$$

where w_i is the log fitness of the ith chromosome in the population, given by $s_a y_i - s_d z_i$ (for y_i, z_i, the advantageous and deleterious mutants, respectively, in genome $i \in N$). $\tilde{y}_j$, $\tilde{z}_j$ are the number of genomes that a particular mutant appears in, divided by the total population size; that is, they are the frequencies of a segregating advantageous or deleterious mutant at locus j (with m segregating loci overall). Each locus has no more then two alleles segregating at any one time in this simulation.

For a population with recombination, a new mutant has a map position attributed to it drawn from a uniform [0, 1] distribution. During reproduction, the position of a crossover is drawn from the same distribution. If one crossover is chosen, allelic states are exchanged at sites where map position exceeds the recombination distance. If two crossovers occur, the states of loci are swapped where the mutant map position lies between the two crossover points. More than two crossovers are unlikely (the probability of more than two occurring is 0.00015, with $L = 0.1$); therefore only up to two exchanges are considered.

Measuring the strength of selection on a modifier: To measure selection on a modifier, a modifier allele is introduced at a frequency of 50% into a population after a burn-in. Introducing the modifier at an intermediate frequency prevents its immediate loss (or fixation), which would otherwise bias our long-term estimate of selection by forcing it to equal zero for all generations following its premature loss (or fixation). After its introduction, a modifier is tracked for 200 generations or until it is fixed or lost. At each generation following its introduction the change in modifier frequency ΔM is noted, and selection on the modifier is estimated using the weak-selection equation $s_M = \Delta M/(p_M q_M)$ (Barton 1995). The value of s_M at the 200th generation is taken as the overall strength of selection acting on the modifier. This is repeated for $5N$ modifiers per burn-in, so a distribution of average selection strengths is developed. This is repeated for 100 burn-ins.

Measuring N_e for asexual populations: To estimate N_e, a neutral, linked locus is inserted into the genome at a random position (*i.e.*, the possibility of it being telomeric or centromeric is allowed). This locus affects a quantitative trait, which has an initial effect of zero. After a burn-in, the effect of this locus is changed in each individual by adding Gaussian noise each generation with a mean of zero and variance V equal to one. It can be shown that the equilibrium variance should be VN_e for such a neutral trait (Lynch and Hill 1986). The simulation is left to run with Gaussian-distributed mutations occurring every generation at the neutral

locus for a further $5N$ generations to reach equilibrium, at which point the variance (and N_e) is measured. Average N_e values from independent burn-ins are calculated to form an overall distribution.

RESULTS

Effects of advantageous mutations on a recombination modifier: We first investigate the dynamics of a recombination modifier in the presence of different types of mutations (deleterious only, advantageous only, and both deleterious and advantageous). As observed by Keightley and Otto (2006), we found that the relative fixation probability of a modifier (u/u^*) rises as N increases for all cases investigated. Also, $u/u^* > 1$ for all simulations, indicating that a recombination modifier is always favored. Full results for all scenarios investigated are provided in supporting information, File S1.

Although advantageous mutants arise in our simulations at a low frequency (the proportion of advantageous mutants is $x = k/s_a$, so for $s_a = 0.01$ only 2.3% of mutations are advantageous), their occurrence still causes a high fixation rate of the modifier, even in the absence of deleterious mutations. For example, with an advantageous mutation rate of $Ux = 0.0115$ and $s_a = 0.01$, the relative fixation probability $u/u^* = 3.58$ for $N = 1000$, which is only slightly lower than that observed with deleterious mutations only and $U = 0.5$. The extent to which beneficial mutations select for recombination is even greater in larger populations (for example, $u/u^* = 47.4$ for $N = 10{,}000$), which is greater than the corresponding value for the deleterious mutations case and $U = 0.5$. By increasing s_a to 0.05 but holding constant the net effect of mutations by decreasing the beneficial mutation rate to $Ux = 0.0023$ (as $x = k/s_a$), recombination is even more favorable (*e.g.*, $u/u^* = 81.9$ for $N = 10{,}000$). So in the absence of deleterious mutations, recombination offers substantial benefits in aiding the fixation of recurrent advantageous mutants across multiple loci, especially in large populations.

Figure 1 compares relative rates of recombination modifier fixation for cases with both advantageous and deleterious mutations present (for $N = 25{,}000$ and $U = 0.1$). We observe that the presence of advantageous mutations alongside deleterious mutations leads to a higher fixation probability of a recombination modifier than if mutations are solely deleterious. The highest u/u^* of 220 occurs for the case of weakest selection against deleterious mutations, *i.e.*, $s_d \approx 1/N$, and strong selection in favor of advantageous mutants, *i.e.*, $s_a = 0.05$. For stronger s_d, increased purifying selection acting against deleterious mutants leads to the loss of a larger fraction of advantageous mutations, reducing the extent to which they can contribute to Hill–Robertson interference (Charlesworth 1994; Peck 1994). Even if there are no advantageous mutations, strong purifying selection means that individuals carry few deleterious

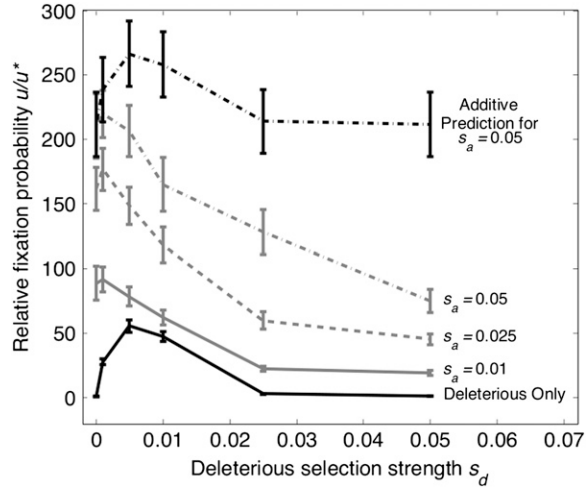


Figure 1.—Relative fixation probability of a recombination modifier u/u^* for $N = 25{,}000$ as a function of the strength of selection acting against deleterious mutants. Mutations are just deleterious or a mixture of deleterious and advantageous with strength $s_a = 0.01$, $s_a = 0.025$, or $s_a = 0.05$. These are compared to the expected u/u^* if both deleterious mutants and advantageous mutants ($s_a = 0.05$) are present and u/u^* is the sum of their independent fixation probabilities (with advantageous mutants only present, $u/u^* \sim 210$ with $s_a = 0.05$). The chromosomal mutation rate in all cases is $U = 0.1$. Bars are 95% confidence intervals here and throughout the article.

mutations in their genome; thus a recombination modifier behaves as more nearly neutral (*e.g.*, $u/u^* = 1.32$ with $N = 25{,}000$, $s_d = 0.05$, and $U = 0.1$, no advantageous mutation).

Such fixation probabilities may depend on the number of linked loci present (Iles *et al.* 2003). Consistent with this, we observe that the fixation probability rises as we increase the number of linked loci from 10 to 50 (Figure S1). However, it appears that fixation probabilities reach a plateau as the number of linked loci approaches 100, indicating that our simulations capture the maximum impact of Hill–Robertson in reducing the efficacy of selection, at least in the population sizes simulated.

With a combination of weak deleterious mutations and strong advantageous mutations, recombination offers a dual advantage, predominantly through the more efficient purging of deleterious mutants [stopping Muller's ratchet (Muller 1964; Felsenstein 1974) and reducing the mutation load (Keightley and Otto 2006)] but also by aiding the fixation of rare advantageous mutants ["Fisher–Muller" hypothesis (Fisher 1930; Muller 1932)]. The increase in modifier fixation with higher s_a is likely to arise because strongly favored mutants are likely to carry along with them many deleterious mutations in the absence of recombination (Peck 1994; Hadany and Feldman 2005) and recombination can free these advantageous mutations from their deleterious backgrounds. In line with this reasoning, Table 1 shows that for $N = 1000$ and $U = 0.5$, re-

TABLE 1

Number of fixed mutants

Case	Strength of deleterious mutations s_d		
	0.01	0.025	0.05
Only deleterious mutations			
Asexual population	1260 (7.07)	823 (4.16)	471 (3.09)
Recombining population	644 (5.99)	314 (3.57)	107 (1.71)
Only advantageous mutations			
Asexual population	195 (2.17)	161 (1.77)	159 (1.43)
Recombining population	378 (2.86)	362 (2.08)	371 (1.95)
Both deleterious and advantageous mutations			
No. of deleterious fixed, asexual	1430 (8.88)	861 (5.09)	476 (3.14)
No. of deleterious fixed, recombining	985 (8.10)	393 (3.63)	125 (2.20)
No. of advantageous fixed, asexual	114 (1.91)	64.6 (1.54)	41.8 (1.27)
No. of advantageous fixed, recombining	269 (1.92)	149 (2.43)	90.1 (1.60)

The average number of mutants that fix over $5N$ generations for $N = 1000$ and $U = 0.5$ is given to three significant values. Cases considered are those where mutations are solely deleterious, advantageous only, or both deleterious and advantageous. Fixations are measured for an asexual population or a population with a constant rate of recombination. Note that if advantageous and deleterious mutants are present, the strength relates to s_d, with $s_a = 0.05$. Values in parentheses are 95% confidence intervals here and throughout the article. When measuring the number of mutants fixed with recombination, the population recombines throughout the burn-in.

combination aids the fixation of advantageous mutants and decreases the fixation rate of deleterious mutants in all cases simulated.

Selection and fixation probabilities of a modifier: We next asked whether the benefits of the modifier brought about by purging deleterious mutants and fixing advantageous mutants are additive; recall that this means that if the modifier has selection advantage s_{Md} if just deleterious mutations occurs at rate U_d, and advantage s_{Ma} if just advantageous mutations occurs at rate U_a, then if both types of mutations occur at a total rate of $U_d + U_a$ the selective advantage is expected to equal $s_{Md} + s_{Ma}$. This is tested by comparing the selection coefficients at the 200th generation after the modifier is introduced for $N = 1000$, $U = 0.5$. Selection on the modifier is measured for three deleterious mutation strengths ($s_d = 0.01$, 0.025, and 0.05), where mutations are solely deleterious at rate U and again where both deleterious and advantageous mutations occur (with $s_a = 0.05$, $U = 0.5$). We also investigate the case where mutations are advantageous only (see points with $s_d = 0$; $s_a = 0.05$), which occurs at the reduced rate $Ux = 0.0115$.

Results of this test are outlined in Figure 2. As with u/u^*, if the number of linked loci under selection is increased, s_M values appear to reach a plateau as the number of loci approaches 100 (Figure S2). Whereas the addition of advantageous mutations enhances fixation of the modifier, the observed values of s_{Mb} (modifier strength when mutations are both advantageous and deleterious) fall short of the additive values, $s_{Md} + s_{Ma}$; in fact $s_{Mb} < s_{Md}$ for all $s_d > 0.01$. s_{Mb} can exceed s_{Md} if the modifier is introduced at a low frequency (<10%) and $s_d \leq 0.1$ (Figure S3); however, s_{Mb} still falls short of the additive prediction. These selection coefficients lie in contrast to the relevant fixation probabilities, u/u^*, as these values increase if advantageous and deleterious mutants are present, compared to the deleterious only case. However, these fixation probabilities also act in a subadditive manner (see also File S1 and Figure 1). This decrease in s_M if advantageous and deleterious mutants are both present seems to verify the hypothesis that extra interference is present if two types of mutations are present together; breaking this down offers an increase. Selection on the modifier also increases with N (Figure S4), consistent with the hypothesis that a recombination modifier is more strongly selected for in larger populations.

Interestingly, increasing s_d increases selection on the modifier, s_M, with or without advantageous mutations, whereas the fixation probability of the modifier u/u^* decreases with $s_d \geq 0.01$ in all simulations (Figure 1). The explanation of this paradoxical result is connected with changes in the effective population sizes N_e and how Hill–Robertson interference affects fixation of the modifier. The fixation probability of a new mutant is determined by its selection strength s and the effective population size N_e according to $u = (1 - \exp(-2sN_e/N))/(1 - \exp(-2sN_e))$ (Kimura 1983). By reducing N_e, Hill–Robertson effects can reduce the fixation probability of a new mutation (the recombination modifier in this case), even if it is more strongly favored (Table 2). With respect to a modifier, having more strongly selected deleterious mutations has a more dramatic impact on reducing N_e than increasing s_M, with the net result that the modifier is less likely to fix.

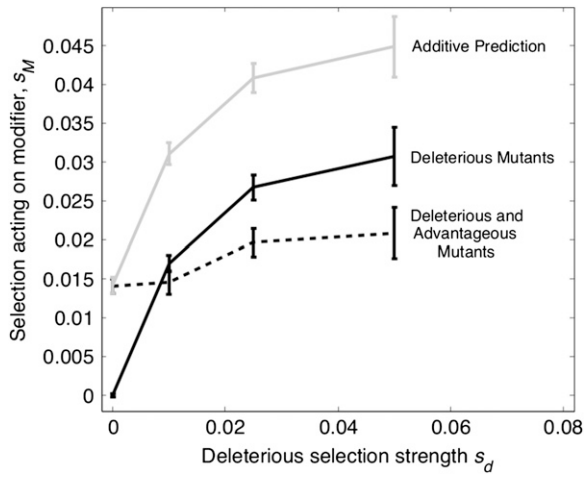


FIGURE 2.—Selective advantage of a modifier, s_M, inferred for cases where mutations are deleterious only and where mutations are deleterious and advantageous. This is compared to the "additive" prediction for the deleterious and advantageous case ($s_M \approx 0.013$ if mutations are solely advantageous, which is the result for the deleterious and advantageous mutant case if $s_d = 0$). $N = 1000$, $U = 0.5$, $s_a = 0.05$ if present.

Whereas Kimura's formula offers accurate estimates of u/u^* if mutations are deleterious, it underestimates fixation probabilities if advantageous mutants are present as well. It appears that selective sweeps alter N_e as mutations rise in frequency, by reducing fitness variance at linked sites (MAYNARD SMITH and HAIGH 1974), violating the assumption that N_e is constant at steady state. This result also suggests that the presence of Hill–Robertson effects can increase the fixation probability of a modifier, relative to that expected at a single locus. This could be due to recombination increasing fitness variance as a modifier rises in frequency (see below), increasing N_e from the value in an asexual population as interference is broken down.

Testing the effectiveness of the diffusion approximation in predicting modifier fixation rates: The previous stochastic simulations are limited in the sense that they can be run only for population sizes that are small, compared to some of the large effective population sizes found in nature. To predict outcomes for larger N, we now investigate diffusion approximations. These predict that the behavior of a new mutant is left unchanged if $N_e\mu$, $N_e s$, and $N_e r$ are constant and small (note that we refer to μ, the per site mutation rate, as opposed to U, the per chromosome mutation rate; $\mu = U/100$ and r is the recombination fraction between individual loci). The diffusion approximation should also hold if $N_e r$ is large and $N_e\mu$ and $N_e s$ are kept small but we do not focus on that situation here. A thorough overview of such work can be found in EWENS (2004) with an example provided by GORDO and CHARLESWORTH (2000). However, the diffusion approximation may not hold for our simulation, since N_e changes with different mutation rates and increases with higher rates of recombination.

TABLE 2

Estimates of s_M and N_e

s_d	Modifier s_M	N_e	Pred. u/u^*	Observed u/u^*
Deleterious mutants only case				
0.01	0.0170 (0.0010)	117.88 (8.41)	4.073	4.518
0.025	0.0268 (0.0016)	73.08 (4.36)	3.989	4.000
0.05	0.0308 (0.0034)	55.14 (2.95)	3.508	3.660
Deleterious and advantageous mutants case				
0.01	0.0146 (0.0015)	97.62 (5.26)	3.021	5.206
0.025	0.0197 (0.0018)	72.48 (3.80)	3.026	4.456
0.05	0.0209 (0.0033)	52.29 (2.60)	2.460	3.900

s_M and N_e are measured for different scenarios investigated ($N = 1000$, $U = 0.5$, s_M plotted in Figure 2), along with predicted fixation rates based on these values using Kimura's formula ("Pred."). These are compared to the fixation rate of the modifier u/u^* obtained from simulations.

We decided to calculate u/u^* for $N\mu = 1$, $Ns = 5$, and $NL = 5$ for all cases of mutation (deleterious only, advantageous only, and both deleterious and advantageous), to determine whether fixation is constant as a function of N. These results are plotted in Figure 3a: the graph shows that u/u^* becomes approximately constant as a function of N, albeit at low values ($\sim$1.25). This suggests that diffusion approximations might be useful as a guide to predict modifier behavior for larger N than is possible to simulate directly. However, as Figure 3b shows, u/u^* increases nonlinearly with N if simulations are run for large $N\mu = 10$, $NL = 1000$, and $Ns = 100$. These parameter values are chosen so that $U = 0.1$, $L = 0.1$, and $s = 0.01$ if $N = 10{,}000$. The nonlinear estimates of u/u^* obtained imply that we cannot extrapolate simulation results unless N is much larger if rates of mutation, recombination, and strength of selection are of these magnitudes (*i.e.*, comparable to parameters observed for Drosophila). The observation that selection on the modifier is not invariant when $N\mu$, NL, and Ns are held constant but large could be either due to a breakdown in the diffusion approximations or due to changes in N_e caused by recombination reducing Hill–Robertson interference.

Effects of variance and linkage disequilibrium on modifier selection: In this section, we investigate how linkage disequilibrium changes with recombination and whether these values relate to u/u^*. Figure 4 compares the genic variance, genetic variance, and variance due to linkage disequilibrium in asexual and recombining populations for two mutational cases (deleterious only and deleterious and advantageous) for $N = 5000$ and $U = 0.1$. In both cases plotted, the genic and genetic variance is unchanged or it increases in the presence of recombination. With deleterious mutations only, the genetic variance is $\sim s_d^2(U/s_d) = Us_d$, the interference-free value of expected variance. If both advantageous and deleterious mutants segregate, the variance in-

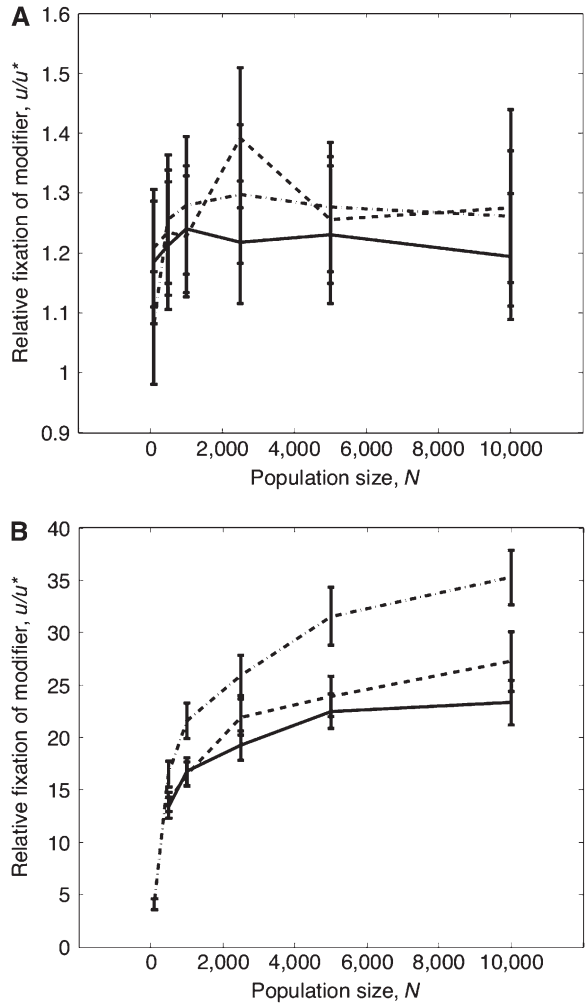


FIGURE 3.—(a) u/u^* as a function of N for fixed $N\mu$, Ns, and NL, which are of $O(1)$ (μ is the per site mutation rate, $\mu = U/100$) if mutations are deleterious (solid line), advantageous (dotted-dashed line) or deleterious and advantageous (dashed line). (b) u/u^* as a function of N with $N\mu$, Ns, and NL fixed but they are no longer $O(1)$.

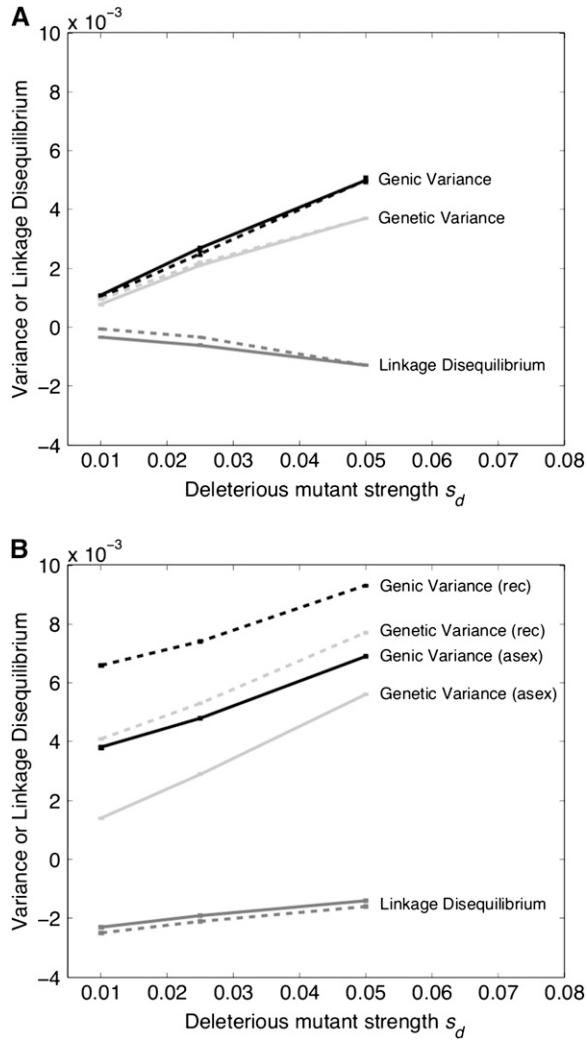


FIGURE 4.—Genic variance, genetic variance, and linkage disequilibrium for an asexual (solid lines) and a recombining (dashed lines) population. (a) The deleterious only case; (b) the advantageous and deleterious mutations case (with $s_a = 0.05$). $N = 5000$, $U = 0.1$.

creases more substantially with recombination compared to the deleterious only case. By Fisher's fundamental theorem of natural selection, this increase in genetic variance should hasten the response to selection and improve the population mean fitness (FISHER 1930; PRICE 1972). Recombination is selected for through association with this rise in fitness.

In Figure 4, a and b, the magnitude of linkage disequilibrium is only slightly different in a recombining population. For $s_d = 0.01$–0.025, where modifier fixation is greatest, the magnitude of linkage disequilibrium increases by ~10-fold if advantageous mutants are present alongside deleterious mutants, compared to the deleterious only case. This signifies a large amount of extra interference being created with the presence of advantageous mutations. However, there is not a one-to-one correspondence between increases in linkage disequilibrium and increases in u/u^*. For example, with $s_d = 0.05$, u/u^* is 11-fold higher in the presence of both advantageous and deleterious mutants than in the case with only deleterious mutants, despite there being little difference in the linkage disequilibrium present.

A possible explanation for this mismatch between the observed level of linkage disequilibrium and the fate of a modifier of recombination is that as recombination breaks down linkage disequilibrium, more advantageous alleles are rescued from poor genetic backgrounds, which increases their chance of establishment and creates extra interference. Due to this, the better predictor for the increase in modifier fixation rate is the increase in genic variance V_g within a population (BARTON and OTTO 2005). For example, with $s_d = 0.01$ genic variance increases by 0.0028 in a recombining population compared to an asexual population. If $s_d = 0.05$, the increase is only by a value of 0.0024. However, the corresponding fixation probability

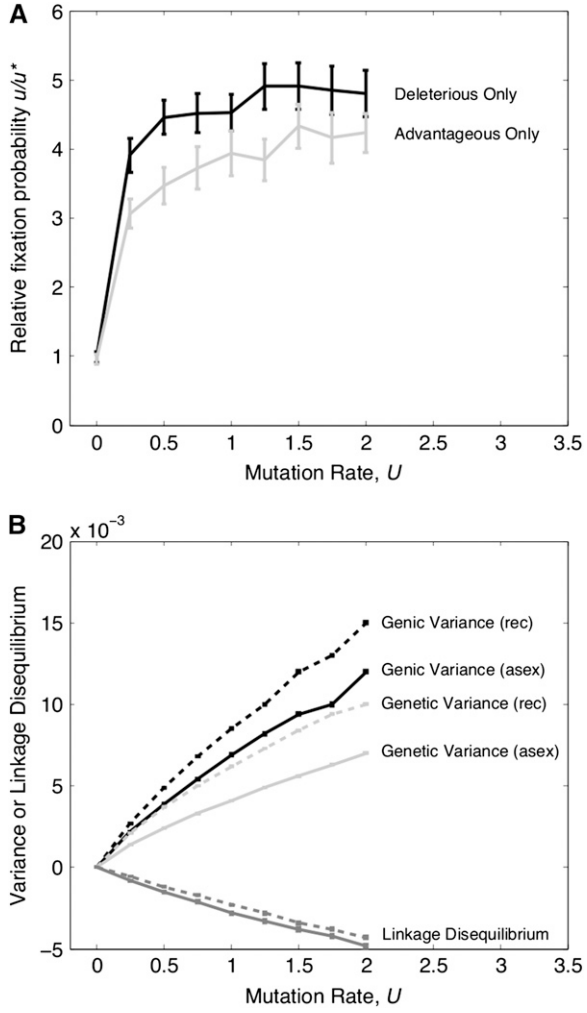


FIGURE 5.—(a) u/u^* as a function of U if mutation is deleterious only and if mutation is advantageous only at rate Ux. $N = 1000$, $s_d = s_a = 0.01$. (b) Genic variance, genetic variance, and linkage disequilibrium as a function of U where mutations are deleterious only, for the same parameters. Populations are asexual (solid lines) or recombining (dashed lines).

u/u^* drops from 29.85 if $s_d = 0.01$ to 14.52 with $s_d = 0.05$. So although the difference in genic variance between an asexual and a recombining population decreases with stronger selection acting against deleterious mutants, the drop is not large enough to predict the steep decline in fixation probability associated with these parameter values.

Does the advantage of recombination continue to rise with *U*? Interestingly, the relative fixation probability of a modifier does not rise linearly with the mutation rate (Figure 5a). This result holds even if the number of linked loci under selection is reduced (Figure S5a). This is unexpected, as one might assume that the fixation probability of a recombination modifier increases with higher U, since more mutants are produced that create extra interference between sites. One reason for this behavior is that as the mutation rate increases, the extent of Hill–Robertson interference also increases, reducing N_e and the spread of a modifier. Thus, even though selection on the modifier, s_M, rises with U (see Table 3 for the deleterious only case), the two effects cancel, leaving the fixation rate of the modifier relatively constant as U increases. s_M values are approximately equal to those shown if there are fewer linked loci under selection (Figure S5b).

TABLE 3

Estimates of s_M and N_e based on values at a neutral locus and predicted/observed u/u^* values as a function of U

U	Modifier s_M	N_e	Predicted u/u^*	Observed u/u^*
0.25	0.0119 (0.00067)	145.79 (4.51)	3.575	3.91
0.50	0.0172 (0.00097)	114.78 (3.28)	4.018	4.46
0.75	0.0213 (0.0010)	103.54 (3.11)	4.455	4.52
1.00	0.0220 (0.0013)	94.83 (2.77)	4.229	4.53
1.25	0.0230 (0.0016)	88.05 (2.61)	4.114	4.91
1.50	0.0240 (0.0017)	83.23 (2.34)	4.062	4.91
1.75	0.0250 (0.0017)	78.77 (2.03)	4.009	4.85
2.00	0.0254 (0.0018)	75.48 (2.12)	3.912	4.80

$N = 1000$, $s_d = 0.01$.

This argument is supported by investigating the underlying genetic and genic variances (Figure 5b). As U increases, the rise in the magnitude of genic variance with recombination becomes larger, indicating greater selection for the modifier. However, the magnitude of linkage disequilibrium also increases due to the presence of more segregating polymorphisms, which will drive down the effective population size. Thus, while one might expect Hill–Robertson effects to select for recombination in direct proportion to the mutation rate, genetic interference is, to a large extent, self-limiting, and we see a strong diminishing returns relationship that tapers off once chromosomewide mutation rates reach $\sim U = 1$.

DISCUSSION

In this article we show that for mutation rates and mean selection strengths that are representative of what is known in Drosophila, the presence of advantageous mutations can lead to substantial selection on a modifier for recombination. As Figure 1 demonstrates, the highest advantages occur if s_d is low and s_a is high. Hence the low rate of adaptive amino acid substitutions observed in Drosophila is capable of aiding the evolution of recombination and can help to account for its widespread occurrence.

That said, the addition of advantageous mutations alongside deleterious mutations increases the fixation of a modifier in a subadditive fashion (File S1 and Figure 1). This demonstrates that whereas there can be a pluralist advantage on recombination in fixing bene-

ficial alleles as well as purging deleterious mutants, the benefit gained from aiding selective sweeps is not as great as we might have expected if assuming that the modifier acts on deleterious mutants and advantageous mutants independently. This arises due to extra interference being created by deleterious mutants causing the loss of beneficial mutants, as highlighted in Table 1.

This study also offers insight into how to best measure the rate of evolution of a recombination modifier in the presence of Hill–Robertson interference. Figure 2 shows that a recombination modifier can be strongly selected for if introduced at 50% frequency; however, due to the high levels of Hill–Robertson interference present (Table 2), it has a fixation rate that is only slightly higher than that of a neutral mutant ($u/u^* = 1.32$). This suggests that measuring the strength of selection acting on a strong modifier can be misleading, since it does not take into account how Hill–Robertson interference impedes the spread of a beneficial mutant. This interference is broken down as a modifier increases in frequency in an initially asexual population, increasing N_e over time. However, if there was already some recombination present in the population, or if the modifier is weak, then N_e would not appreciably change, so s_M might offer accurate insight into the fate of a recombination modifier in these cases.

We also demonstrate that whereas it is theoretically possible to extrapolate fixation values of the modifier for larger N from fixation rates for small N using diffusion models, these assumptions will hold best if the values Ns, $N\mu$, and Nr are of $O(1)$, which predict small rates of modifier fixation. Using larger N and parameter values, diffusion approximations break down, and thus we have to resort to full simulation.

By examining the genic and genetic variance in the simulations (Figure 4), we observe that negative linkage disequilibrium is created, which is indicative of Hill–Robertson interference (Hill and Robertson 1966). Recombination increases genetic variance in fitness within a population, in line with existing theory on the evolution of a recombination modifier in the presence of drift (Barton and Otto 2005). These results, however, highlight an important point that even though linkage disequilibrium is indicative of interference, the magnitude of it does not determine the change in frequency of a modifier (Barton 1995; Barton and Otto 2005). This is exemplified if $s_d = 0.05$, where linkage disequilibrium values are similar in the presence and absence of beneficial mutants, yet modifiers of recombination are more strongly favored in the latter case. This is because a particular level of genomewide linkage disequilibrium (measured by $V_A - V_g$) can arise when there are many segregating deleterious mutations (in which case, advantageous mutations are unlikely to fix) or when there are few segregating deleterious mutations and more advantageous mutations are able to establish.

Overall, this study demonstrates how beneficial mutations provide strong selection for a recombination modifier. However, there are a few caveats associated with the parameters used in this study, which should be investigated further to determine the full extent of the evolution of a recombination modifier.

Values of s_a used in these simulations are higher than those inferred for amino acid substitutions in Drosophila, to prevent drift overwhelming mutations for small population sizes that we investigate. Using smaller values of s_a will certainly reduce the fixation probability of the modifier, at least for the population sizes investigated in these simulations.

Dominance in diploid deleterious mutations is also not considered here. Roze (2009) showed how, if deleterious mutants are highly recessive, recombination is selected against, because breaking apart multilocus heterozygosity incurs fitness disadvantages, especially if selection on the mutations is weak. Future work should investigate how the presence of advantageous mutations affects this result, although such a study is likely to strongly rely on the dominance of beneficial mutations, which is only poorly known.

The strength of selection on adaptive mutations also depends on the organism under observation and the state of its environment. Relative fitness differences in Drosophila can be reduced if their populations are dense or if there is a lack of available food (Kondrashov and Houle 1994). On the other hand, in bacteria and viruses, advantageous mutants with larger fitness effects have been observed in stressed environments [on the order of $s = 5$ (Barrett *et al.* 2006) or even $s = 12$ (Bull *et al* (2000)]. Bearing all this in mind, the effects on a modifier over a larger range of selection parameters should be investigated.

These results also offer predictions as to when recombination can evolve. If an organism moves to a new environment to which it is maladapted, our model predicts that higher rates of recombination are more likely to arise in this new environment. Furthermore, if background deleterious mutations are frequent, recombination has an extra advantage in aiding purifying selection and is even more likely to evolve than in the presence of beneficial mutations alone. Such a scenario was discussed by Hadany and Feldman (2005) and could explain why recombination is more likely to occur in new, stressed environments (Abdullah and Borts 2001; Grishkan *et al.* 2003).

Finally, we did not investigate whether the advantages to a recombination modifier in the presence of advantageous mutants transfer over to a sex modifier. This requires an adjusted model to account for the costs of sex (Maynard Smith 1978) and to ensure that excessive inbreeding is avoided (if rare sexuals can mate only with other sexuals). This is a well-known problem with regard to the evolution of sex (see, for example, Peck 1993) and we will investigate this in a future research article.

We thank Nick Barton, Brian Charlesworth, Nick Colegrave, Denis Roze, and an anonymous reviewer for helpful comments on the manuscript. M.H. is funded by a Biotechnology and Biological Sciences Research Council (BBSRC) studentship and a University of British Columbia Graduate Student International Research Mobility Award; S.P.O. is funded by the Natural Sciences and Engineering Research Council of Canada; P.D.K. acknowledges support from the BBSRC. This work made use of the resources provided by the Edinburgh Compute and Data Facility (ECDF) (http://www.ecdf.ed.ac.uk). The ECDF is partially supported by the e-Science Data, Information and Knowledge Transformation initiative (http://www.edikt.org.uk).

LITERATURE CITED

Abdullah, M. F. F., and R. H. Borts, 2001 Meiotic recombination frequencies are affected by nutritional states in *Saccharomyces cerevisiae*. Proc. Natl. Acad. Sci. USA **98:** 14524–14529.

Andolfatto, P., 2005 Adaptive evolution of non-coding DNA in *Drosophila*. Nature **437:** 1149–1152.

Andolfatto, P., 2007 Hitchhiking effects of recurrent beneficial amino acid substitutions in the *Drosophila melanogaster* genome. Genome Res. **17:** 1755–1762.

Barrett, R. D. H., R. Craig MacLean and G. Bell, 2006 Mutations of intermediate effect are responsible for adaptation in evolving *Pseudomonas fluorescens* populations. Biol. Lett. **2:** 236–238.

Barton, N. H., 1995 A general model for the evolution of recombination. Genet. Res. **65:** 123–144.

Barton, N. H., 2009 Why sex and recombination? Cold Spring Harbor Symp. Quant. Biol. (in press).

Barton, N. H., and B. Charlesworth, 1998 Why sex and recombination? Science **281:** 1986–1990.

Barton, N. H., and S. P. Otto, 2005 Evolution of recombination due to random drift. Genetics **169:** 2353–2370.

Betancourt, A. J., J. J. Welch and B. Charlesworth, 2009 Reduced effectiveness of selection caused by a lack of recombination. Curr. Biol. **19:** 655–660.

Bierne, N., and A. Eyre-Walker, 2004 The genomic rate of adaptive amino acid substitution in *Drosophila*. Mol. Biol. Evol. **21:** 1350–1360.

Bull, J. J., M. R. Badgett and H. A. Wichman, 2000 Big-benefit mutations in a bacteriophage inhibited with heat. Mol. Biol. Evol. **17:** 942–950.

Bulmer, M. G., 1976 The effect of selection on genetic variability: a simulation study. Genet. Res. **28:** 101–117.

Bulmer, M. G., 1980 *The Mathematical Theory of Quantitative Genetics*. Clarendon Press, Oxford.

Burt, A., 2000 Sex, recombination, and the efficacy of selection—was Weismann right? Evolution **54:** 337–351.

Charlesworth, B., 1994 The effect of background selection against deleterious mutations on weakly selected, linked variants. Genet. Res. **63:** 213–227.

Charlesworth, B., and D. Charlesworth, 1975 An experiment on recombination load in *Drosophila melanogaster*. Genet. Res. **25:** 267–274.

Charlesworth, B., M. T. Morgan and D. Charlesworth, 1993 The effect of deleterious mutations on neutral molecular variation. Genetics **134:** 1289–1303.

Charlesworth, B., A. J. Betancourt, V. B. Kaiser and I. Gordo, 2009 Genetic recombination and molecular evolution. Cold Spring Harbor Symp. Quant. Biol. (in press).

Chimpanzee Sequencing and Analysis Consortium, 2005 Initial sequence of the chimpanzee genome and comparison with the human genome. Nature **437:** 69–87.

Colegrave, N., 2002 Sex releases the speed limit on evolution. Nature **420:** 664–666.

Comeron, J. M., A. Williford and R. M. Kliman, 2008 The Hill-Robertson effect: evolutionary consequences of weak selection and linkage in finite populations. Heredity **100:** 19–31.

Cooper, T. F., 2007 Recombination speeds adaptation by reducing competition between beneficial mutations in populations of *Escherichia coli*. PLoS Biol. **5:** e225.

Crow, J. F., 1970 Genetic loads and the cost of natural selection, pp. 128–177 in *Mathematical Topics in Population Genetics* (Biomathematics, Vol. 1), edited by K.-I. Kojima. Springer-Verlag, Berlin.

de Oliveira, V. M., J. K. da Silva and P. R. A. Campos, 2008 Epistasis and the selective advantage of sex and recombination. Phys. Rev. E **78:** 031905.

Ewens, W. J., 2004 *Mathematical Population Genetics: 1. Theoretical Introduction* (Interdisciplinary Applied Mathematics, Vol. 27, Ed. 2). Springer, New York.

Eyre-Walker, A., 2006 The genomic rate of adaptive evolution. Trends Ecol. Evol. **21:** 569–575.

Eyre-Walker, A., and P. D. Keightley, 2009 Estimating the rate of adaptive molecular evolution in the presence of slightly deleterious mutations and population size change. Mol. Biol. Evol. **26:** 2097–2108.

Felsenstein, J., 1974 The evolutionary advantage of recombination. Genetics **78:** 737–756.

Fisher, R. A., 1930 *The Genetical Theory of Natural Selection*. Clarendon Press, Oxford.

García-Dorado, A., C. López-Fanjul and A. Caballero, 1999 Properties of spontaneous mutations affecting quantitative traits. Genet. Res. **74:** 341–350.

Goddard, M. R., H. C. J. Godfray and A. Burt, 2005 Sex increases the efficacy of natural selection in experimental yeast populations. Nature **434:** 636–640.

Gordo, I., and B. Charlesworth, 2000 The degeneration of asexual haploid populations and the speed of Muller's ratchet. Genetics **154:** 1379–1387.

Grishkan, I., A. B. Korol, E. Nevo and S. P. Wasser, 2003 Ecological stress and sex evolution in soil microfungi. Proc. R. Soc. B **270:** 13–18.

Haag-Liautard, C., M. Dorris, X. Maside, S. Macaskill, D. L. Halligan *et al.*, 2007 Direct estimation of per nucleotide and genomic deleterious mutation rates in *Drosophila*. Nature **445:** 82–85.

Hadany, L., and J. M. Comeron, 2008 Why are sex and recombination so common? Ann. NY Acad. Sci. **1133:** 26–43.

Hadany, L., and M. W. Feldman, 2005 Evolutionary traction: the cost of adaptation and the evolution of sex. J. Evol. Biol. **18:** 309–314.

Halligan, D. L., and P. D. Keightley, 2006 Ubiquitous selective constraints in the *Drosophila* genome revealed by a genome-wide interspecies comparison. Genome Res. **16:** 875–884.

Hill, W. G., and A. Robertson, 1966 The effect of linkage on limits to artificial selection. Genet. Res. **8:** 269–294.

Iles, M. M., K. Walters and C. Cannings, 2003 Recombination can evolve in large finite populations given selection on sufficient loci. Genetics **165:** 2249–2258.

Jensen, J. D., K. R. Thornton and P. Andolfatto, 2008 An approximate Bayesian estimator suggests strong, recurrent selective sweeps in *Drosophila*. PLoS Genet. **4:** e1000198.

Keightley, P. D., and W. G. Hill, 1987 Directional selection and variation in finite populations. Genetics **117:** 573–582.

Keightley, P. D., and S. P. Otto, 2006 Interference among deleterious mutations favours sex and recombination in finite populations. Nature **443:** 89–92.

Keightley, P. D., U. Trivedi, M. Thomson, F. Oliver, S. Kumar *et al.*, 2009 Analysis of the genome sequences of three *Drosophila melanogaster* spontaneous mutation accumulation lines. Genome Res. **19:** 1195–1201.

Kimura, M., 1983 *The Neutral Theory of Molecular Evolution*. Cambridge University Press, Cambridge, UK.

Kondrashov, A. S., and D. Houle, 1994 Genotype-environment interactions and the estimation of the genomic mutation rate in *Drosophila melanogaster*. Proc. R. Soc. Lond. Ser. B Biol. Sci. **258:** 221–227.

Loewe, L., and B. Charlesworth, 2006 Inferring the distribution of mutational effects on fitness in *Drosophila*. Biol. Lett. **2:** 426–430.

Lynch, M., and W. G. Hill, 1986 Phenotypic evolution by neutral mutation. Evolution **40:** 915–935.

Lynch, M., and B. Walsh, 1988 *Genetics and Analysis of Quantitative Traits*. Sinauer Associates, Sunderland, MA.

Malmberg, R. L., 1977 The evolution of epistasis and the advantage of recombination in populations of bacteriophage T4. Genetics **86:** 607–621.

Martin, G., S. P. Otto and T. Lenormand, 2006 Selection for recombination in structured populations. Genetics **172:** 593–609.
Maynard Smith, J., 1978 *The Evolution of Sex*. Cambridge University Press, Cambridge, UK/New York.
Maynard Smith, J., 1988 Selection for recombination in a polygenic model—the mechanism. Genet. Res. **51:** 59–63.
Maynard Smith, J., and J. Haigh, 1974 The hitch-hiking effect of a favourable gene. Genet. Res. **23:** 23–35.
Morran, L. T., M. D. Parmenter and P. C. Phillips, 2009 Mutation load and rapid adaptation favour outcrossing over self-fertilization. Nature **462:** 350–352.
Muller, H. J., 1932 Some genetic aspects of sex. Am. Nat. **66:** 118–138.
Muller, H. J., 1964 The relation of recombination to mutational advance. Mutat. Res. **1:** 2–9.
Otto, S. P., 2009 The evolutionary enigma of sex. Am. Nat. **174:** S1–S14.
Otto, S. P., and N. H. Barton, 1997 The evolution of recombination: removing the limits to natural selection. Genetics **147:** 879–906.
Otto, S. P., and N. H. Barton, 2001 Selection for recombination in small populations. Evolution **55:** 1921–1931.
Otto, S. P., and M. W. Feldman, 1997 Deleterious mutations, variable epistatic interactions, and the evolution of recombination. Theor. Popul. Biol. **51:** 134–147.
Peck, J. R., 1993 Frequency-dependent selection, beneficial mutations, and the evolution of sex. Proc. R. Soc. B **254:** 87–92.
Peck, J. R., 1994 A ruby in the rubbish: beneficial mutations, deleterious mutations and the evolution of sex. Genetics **137:** 597–606.
Poon, A., and L. Chao, 2004 Drift increases the advantage of sex in RNA bacteriophage ϕ6. Genetics **166:** 19–24.
Presgraves, D. C., 2005 Recombination enhances protein adaptation in *Drosophila melanogaster*. Curr. Biol. **15:** 1651–1656.
Price, G. R., 1972 Fisher's 'fundamental theorem' made clear. Ann. Hum. Genet. **36:** 129–140.
Rice, W. R., and A. K. Chippindale, 2001 Sexual recombination and the power of natural selection. Science **294:** 555–559.
Robertson, A., 1961 Inbreeding in artificial selection programmes. Genet. Res. **2:** 189–194.
Roze, D., 2009 Diploidy, population structure, and the evolution of recombination. Am. Nat. **174:** S79–S94.
Roze, D., and N. H. Barton, 2006 The Hill-Robertson effect and the evolution of recombination. Genetics **173:** 1793–1811.
Sella, G., D. A. Petrov, M. Przeworski and P. Andolfatto, 2009 Pervasive natural selection in the *Drosophila* genome? PLoS Genet. **5:** e1000495.
Smith, N. G. C., and A. Eyre-Walker, 2002 Adaptive protein evolution in *Drosophila*. Nature **415:** 1022–1024.
Weismann, A., 1887 On the signification of the polar globules. Nature **36:** 607–609.
West, S., C. Lively and A. Read, 1999 A pluralist approach to sex and recombination. J. Evol. Biol. **12:** 1003–1012.
Wright, S. I., and P. Andolfatto, 2008 The impact of natural selection on the genome: emerging patterns in *Drosophila* and *Arabidopsis*. Annu. Rev. Ecol. Syst. **39:** 193–213.
Zhang, L., and W.-H. Li, 2005 Human SNPs reveal no evidence of frequent positive selection. Mol. Biol. Evol. **22:** 2504–2507.

Communicating editor: M. K. Uyenoyama

GENETICS

Supporting Information
http://www.genetics.org/cgi/content/full/genetics.109.112920/DC1

The Role of Advantageous Mutations in Enhancing the Evolution of a Recombination Modifier

Matthew Hartfield, Sarah P. Otto and Peter D. Keightley

DOI: 10.1534/genetics.109.112920

FILE S1

Full List of Modifier Fixation Results

The following section outlines all results collected for the fixation probability of a recombination modifier, u/u^*. Data was collected for N = 100, 500, 1000, 5000, 10,000 and 25,000; $s_a = s_d = 0.01$, 0.025 and 0.05; and $U = 0.1$ and 0.5. Data was also collected with $s_d = 0$, 0.001 and 0.005 if $N \geq 10,000$ and $U = 0.1$.

There are three cases; mutations are deleterious only, advantageous only, or deleterious and advantageous. Results will be presented in this order. Bracketed values are 95% confidence intervals here and throughout the appendix.

S.1.1 u/u^* for deleterious mutations only present

u/u^* as N increases, $U = 0.1$			
N	Deleterious mutations strength s_d		
	0.01	0.025	0.05
100	1.084 (0.087)	1.186 (0.099)	1.158 (0.090)
500	2.068 (0.147)	2.220 (0.182)	1.570 (0.141)
1000	3.147 (0.314)	2.912 (0.258)	1.550 (0.182)
5000	11.79 (0.797)	6.852 (0.609)	1.274 (0.109)
10000	23.34 (2.117)	6.646 (0.901)	1.260 (0.095)
25000	47.55 (3.639)	3.333 (0.302)	1.315 (0.103)
N	Deleterious mutations strength s_d		
	0.000	0.001	0.005
10000	1.060 (0.097)	9.988 (0.645)	22.30 (1.652)
25000	1.026 (0.146)	28.00 (2.226)	55.72 (4.887)
$U = 0.5$			
100	1.174 (0.096)	1.280 (0.100)	1.392 (0.116)
500	2.550 (0.186)	2.502 (0.225)	2.686 (0.220)
1000	4.518 (0.379)	4.000 (0.342)	3.664 (0.312)
5000	16.78 (1.737)	11.18 (2.095)	9.628 (1.102)
10000	31.73 (3.165)	23.59 (2.781)	13.76 (1.242)
25000	70.46 (7.854)	47.16 (4.990)	23.11 (2.620)

S.1.2 u/u^* for advantageous mutations only present

Note that in this case, the actual mutation rate is dependent on the selection strength s_a of the mutant, using the function $Ux = Uk/s_a$ as we are interested in the modifier fixation probability if advantageous mutants occur at the rate they would occur if arising alongside deleterious mutations. This enables us to test whether the selection strength of the modifier is additive compared to the advantageous only and deleterious only case.

u/u^* as N increases, mutation rate $= Uk/s_a = 0.1 \cdot 0.00023/s_a$			
N	Advantageous mutations strength s_a		
	0.01	0.025	0.05
100	1.030 (0.090)	1.074 (0.094)	1.072 (0.088)
500	1.390 (0.121)	1.866 (0.152)	2.274 (0.230)
1000	2.338 (0.173)	3.742 (0.326)	4.112 (0.414)
5000	16.33 (1.345)	25.36 (2.143)	34.11 (3.502)
10000	35.29 (2.625)	58.98 (5.651)	77.56 (7.671)
25000	88.88 (13.20)	161.77 (16.48)	210.74 (24.98)
Mutation rate $= Uk/s_a = 0.5 \cdot 0.00023/s_a$			
100	1.022 (0.075)	1.066 (0.088)	1.122 (0.103)
500	1.870 (0.121)	2.706 (0.225)	3.000 (0.270)
1000	3.576 (0.236)	4.598 (0.425)	6.442 (0.578)
5000	20.84 (1.780)	29.14 (2.623)	35.50 (4.118)
10000	47.44 (4.000)	68.76 (6.968)	81.92 (8.656)
25000	115.62 (12.09)	152.87 (15.99)	184.01 (18.30)

S.1.3 u/u^* for advantageous and deleterious mutations present

u/u^* for $N = 100$, $U = 0.1$			
s_a	Deleterious mutations strength s_d		
	0.01	0.025	0.05
0.01	1.076 (0.084)	1.124 (0.089)	1.188 (0.083)
0.025	1.060 (0.096)	1.114 (0.093)	1.164 (0.095)
0.05	1.092 (0.090)	1.198 (0.091)	1.240 (0.100)
$U = 0.5$			
0.01	1.212 (0.098)	1.226 (0.098)	1.238 (0.112)
0.025	1.164 (0.090)	1.342 (0.120)	1.260 (0.108)
0.05	1.202 (0.108)	1.206 (0.108)	1.330 (0.093)
u/u^* for $N = 500$, $U = 0.1$			
s_a	Deleterious mutations strength s_d		
	0.01	0.025	0.05
0.01	1.882 (0.148)	1.952 (0.164)	1.530 (0.143)
0.025	2.166 (0.174)	2.192 (0.169)	1.772 (0.160)
0.05	2.538 (0.228)	2.214 (0.177)	1.952 (0.176)
$U = 0.5$			
0.01	2.510 (0.190)	2.744 (0.209)	2.716 (0.216)
0.025	2.758 (0.218)	2.824 (0.206)	2.750 (0.242)
0.05	3.222 (0.250)	2.698 (0.220)	2.620 (0.222)
u/u^* for $N = 1000$, $U = 0.1$			
s_a	Deleterious mutations strength s_d		
	0.01	0.025	0.05
0.01	3.326 (0.236)	2.902 (0.203)	1.772 (0.171)
0.025	3.642 (0.244)	3.510 (0.294)	2.128 (0.178)
0.05	4.706 (0.377)	3.666 (0.344)	2.770 (0.254)

$U = 0.5$			
0.01	4.656 (0.419)	4.718 (0.374)	4.120 (0.430)
0.025	4.596 (0.362)	4.318 (0.385)	4.028 (0.322)
0.05	5.206 (0.445)	4.456 (0.436)	3.900 (0.321)

u/u^* for $N = 5000$, $U = 0.1$			
s_a	Deleterious mutations strength s_d		
	0.01	0.025	0.05
0.01	12.41 (1.139)	7.934 (0.708)	4.010 (0.253)
0.025	19.03 (1.395)	12.09 (1.155)	7.172 (0.666)
0.05	29.85 (3.000)	21.33 (2.153)	14.52 (1.586)

$U = 0.5$			
0.01	17.28 (1.540)	15.29 (1.562)	9.366 (0.822)
0.025	22.56 (2.514)	14.00 (1.388)	9.398 (0.874)
0.05	25.23 (2.571)	15.62 (1.752)	10.52 (1.046)

u/u^* for $N = 10{,}000$, $U = 0.1$			
s_a	Deleterious mutations strength s_d		
	0.01	0.025	0.05
0.01	27.27 (2.824)	12.75 (1.201)	8.850 (0.594)
0.025	42.33 (4.377)	22.71 (2.166)	14.97 (1.391)
0.05	61.60 (6.219)	41.26 (4.227)	31.60 (3.915)

s_a	Deleterious mutations strength s_d		
	0.000	0.001	0.005
0.01	35.29 (2.625)	36.06 (2.392)	34.95 (2.935)
0.025	58.98 (5.651)	60.37 (5.802)	52.04 (4.535)
0.05	77.56 (7.671)	81.38 (7.505)	74.36 (6.874)

$U = 0.5$			
0.01	34.79 (3.341)	26.45 (2.919)	15.20 (2.186)
0.025	36.96 (3.835)	26.41 (3.492)	14.82 (1.593)
0.05	55.01 (6.306)	27.65 (3.137)	15.32 (1.408)

u/u^* for N = 25,000, U = 0.1			
s_a	Deleterious mutations strength s_d		
	0.01	0.025	0.05
0.01	62.50 (5.765)	22.74 (1.799)	19.25 (1.898)
0.025	118.38 (10.76)	63.97 (6.686)	48.40 (4.428)
0.05	204.15 (20.05)	143.50 (17.59)	84.48 (9.614)
s_a	Deleterious mutations strength s_d		
	0.000	0.001	0.005
0.01	88.88 (13.20)	91.94 (9.373)	78.62 (7.295)
0.025	161.77 (16.48)	177.02 (16.20)	148.88 (14.45)
0.05	210.74 (24.98)	219.84 (18.43)	206.44 (19.85)

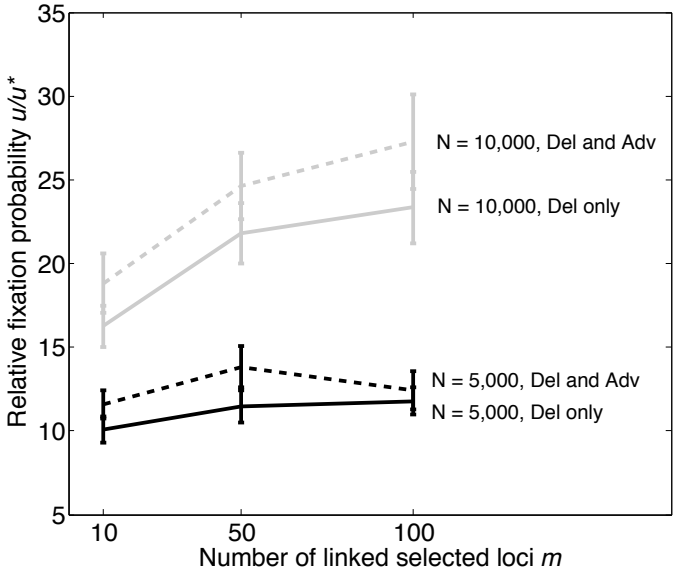
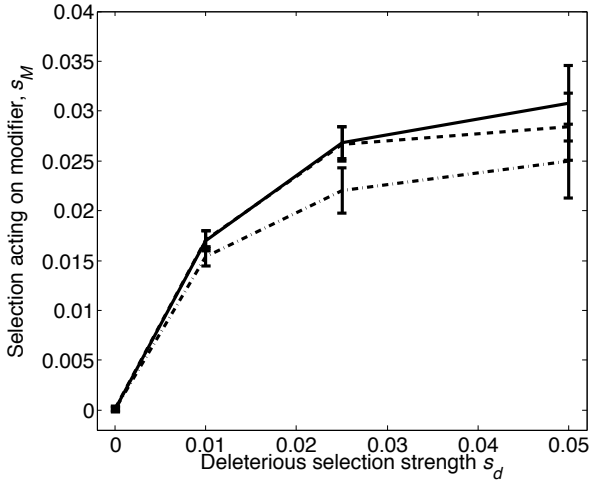
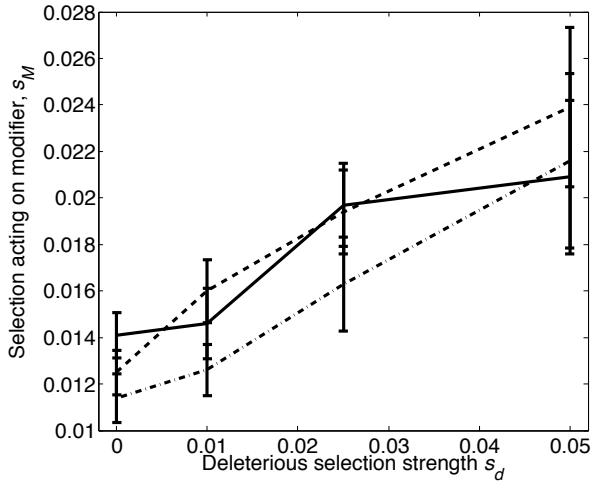


FIGURE S1.—Relative fixation probability of a recombination modifier u/u^* as a function of the number of linked loci under selection. Mutations are just deleterious, or a mixture of deleterious and advantageous with strength $s_a = 0.05$. $U = 0.1$ in all cases simulated.

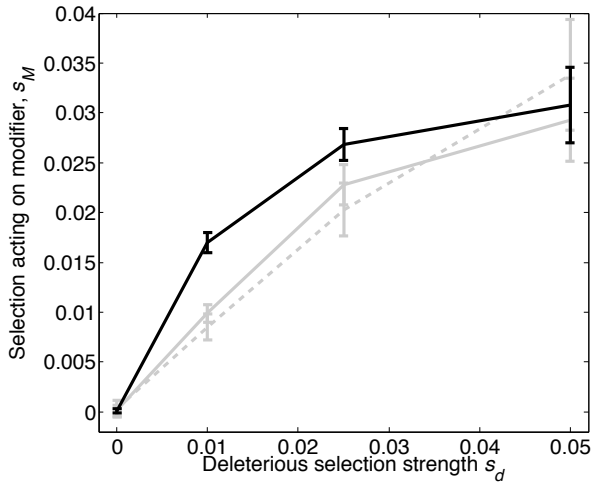


(a)

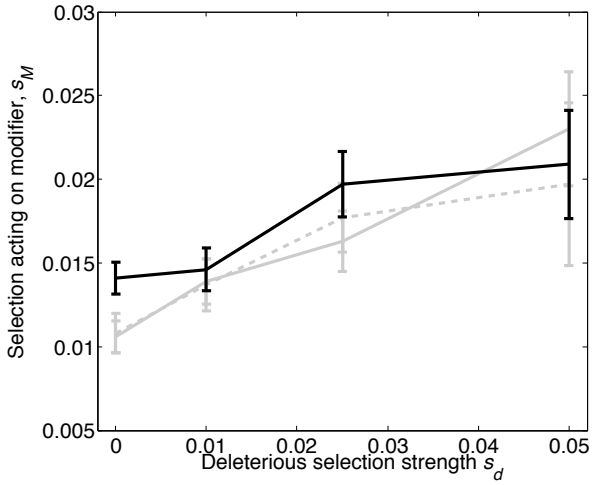


(b)

FIGURE S2.—Selective advantage of a modifier, s_M, inferred for cases where mutations are deleterious only (a) and where mutations are deleterious and advantageous (b). The number of linked loci under selection is either 10 (dot-dashed line), 50 (dashed line) or 100 (solid line). $N = 1000$, $U = 0.5$, $s_a = 0.05$ if present.



(a)



(b)

FIGURE S3.—Selective advantage of a modifier, s_M, inferred for cases where mutations are deleterious only (a) and where mutations are deleterious and advantageous (b). Modifier is introduced at a frequency of 5% (gray dashed line), 10% (gray solid line) or 50% (black line). $N = 1000$, $U = 0.5$, $s_a = 0.05$ if present.

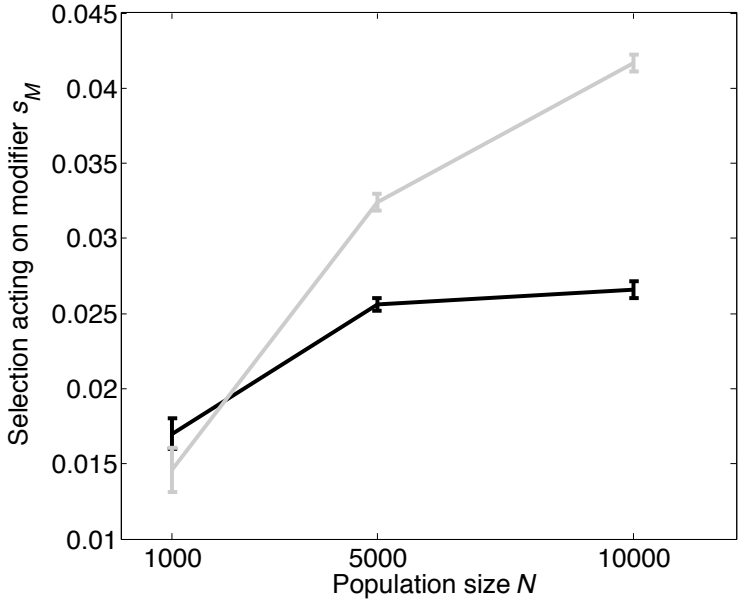


FIGURE S4.—Selective advantage of a modifier, s_M as a function of population size N. Mutations are deleterious only (black line) and where mutations are deleterious and advantageous (gray line). In all cases $U = 0.5$, $s_a = 0.05$ if present.

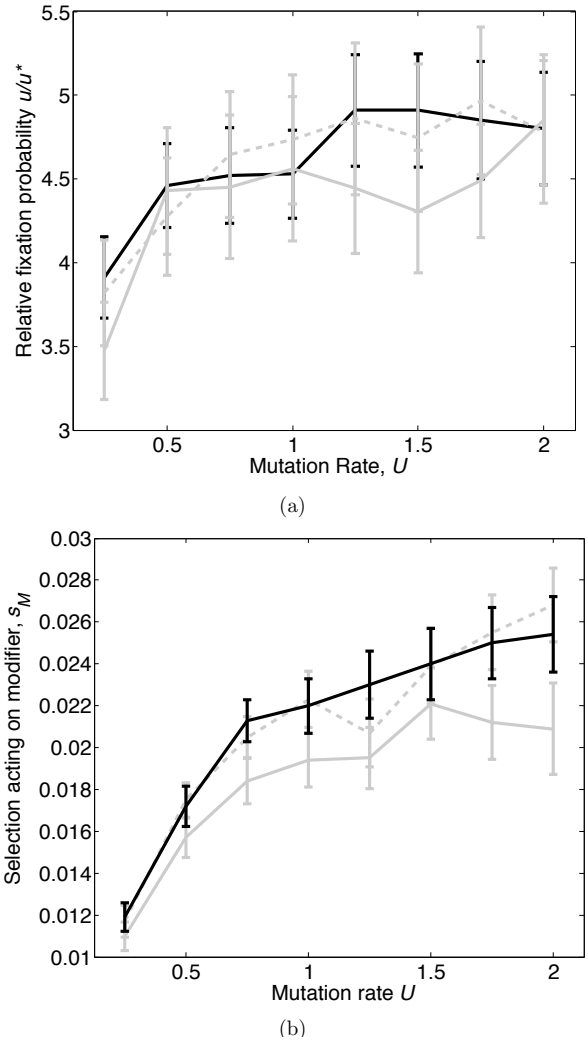


FIGURE S5.—(a) u/u^* and (b) s_M as a function of U if mutation is deleterious only, where there are 10 loci (gray solid line), 50 loci (grey dashed line), or 100 loci (black line) under selection. $N = 1000$, $s_d = 0.01$.

doi:10.1111/j.1558-5646.2011.01311.x

RECOMBINATION AND HITCHHIKING OF DELETERIOUS ALLELES

Matthew Hartfield[1] and Sarah P. Otto[2,3]

[1]*Institute of Evolutionary Biology, School of Biological Sciences, University of Edinburgh, Edinburgh EH9 3JT, United Kingdom*

[2]*Department of Zoology, University of British Columbia, Vancouver, British Columbia V6T 1Z4, Canada*

[3]*E-mail: otto@zoology.ubc.ca*

Received December 17, 2010
Accepted March 18, 2011

When new advantageous alleles arise and spread within a population, deleterious alleles at neighboring loci can hitchhike alongside them and spread to fixation in areas of low recombination, introducing a fixed mutation load. We use branching processes and diffusion equations to calculate the probability that a deleterious allele hitchhikes and fixes alongside an advantageous mutant. As expected, the probability of fixation of a deleterious hitchhiker rises with the selective advantage of the sweeping allele and declines with the selective disadvantage of the deleterious hitchhiker. We then use computer simulations of a genome with an infinite number of loci to investigate the increase in load after an advantageous mutant is introduced. We show that the appearance of advantageous alleles on genetic backgrounds loaded with deleterious alleles has two potential effects: it can fix deleterious alleles, and it can facilitate the persistence of recombinant lineages that happen to occur. The latter is expected to reduce the signals of selection in the surrounding region. We consider these results in light of human genetic data to infer how likely it is that such deleterious hitchhikers have occurred in our recent evolutionary past.

KEY WORDS: **Deleterious mutations, diffusion equations, genetic hitchhiking, Hill–Robertson effects, selective sweep.**

The first generation of evolutionary models of advantageous alleles focused on the dynamics of single selected loci in isolation from surrounding sites (Haldane, 1927; Fisher, 1930). Hill and Robertson (1966) demonstrated, however, that selection acting at one site in finite populations interferes with the efficacy of selection at surrounding sites, hampering the spread of neighboring beneficial alleles, even in the absence of fitness interactions among the sites. As pointed out by Hill and Robertson (1966), Charlesworth et al. (1993), and generalized by Rice (1999), selection on linked sites reduces the effective number of lineages contributing to future generations to those lineages with the highest fitness. Such genetic bottlenecks increase the power of drift relative to selection, such that advantageous alleles are less likely to spread and they spread more slowly than predicted by their direct effects on fitness. In a general analysis by Barton (1995), Hill–Robertson interference was shown to reduce the fixation probability of beneficial alleles linked to other selected sites. Breaking down interference among selected loci has also been shown to favor increased rates of sex and recombination (Otto and Barton, 1997; Barton and Otto, 2005; Roze and Barton, 2006).

In addition to affecting neighboring loci under selection, Maynard Smith and Haigh (1974) showed that the dynamics of a single selected locus impacts surrounding neutral loci. In particular, an advantageous allele sweeping through a population reduces, on average, the genetic variance around the site of a sweep (see also Thomson 1977). This phenomenon provides a mechanism for detecting regions experiencing selection, forming the basis for the Hudson-Kreitman-Aguade (HKA) test (Hudson et al., 1987), for example.

Relatively little attention has been paid, however, to the effect that selection on neighboring sites might have on the net fitness change associated with the fixation of a focal beneficial allele and

on the patterns of variation at surrounding selected sites (see, e.g., Yu and Etheridge (2010) regarding beneficial alleles segregating in the background, and Hadany and Feldman (2005) regarding deleterious alleles in the background). In this article, we consider a focal site carrying a new beneficial allele in the presence of neighboring sites subject to deleterious mutations. We calculate the chance that a linked deleterious allele hitchhikes to fixation along with the beneficial allele, as a function of the rate of recombination between them, and describe the implications for patterns of variation expected within the region of a selective sweep. This work builds upon a recent simulation study by Hadany and Feldman (2005), as well as complementary analytical work for asexual organisms (Johnson and Barton, 2002; Bachtrog and Gordo, 2004; Yu and Etheridge, 2008; Yu et al., 2010). Specifically, Hadany and Feldman (2005) demonstrated that beneficial alleles sweeping to fixation in a purely asexual population often carry along linked deleterious alleles. The fixation of deleterious alleles by hitchhiking generates a fixed mutation load that must await a future adaptive sweep by a back or compensatory mutation in order for it to be erased. Our work provides an analytical prediction of the probability of such undesirable hitchhikers, allowing for arbitrary rates of recombination between the sites under selection.

Empirical Background—Recent studies of amino-acid substitution data suggest that advantageous mutants are present at higher rates then previously assumed. Although precise values remain a matter of debate (Eyre-Walker, 2006), Bierne and Eyre-Walker (2004) estimated that approximately 45% of amino acid substitutions are adaptive in *Drosophila melanogaster*, equating to one substitution, on average, every 450 generations. Later studies have found that between 30 and 60% of substitutions in *D. melanogaster* coding and noncoding regions are adaptive (Andolfatto, 2005; Obbard et al., 2009; Andolfatto, 2007; Shapiro et al., 2007), highlighting the prevalence of beneficial mutation. Similar values have been observed in the wild mouse *Mus musculus castaneus* (Halligan et al., 2010). In hominids this rate tends to be lower; Boyko et al. (2008) and Eyre-Walker and Keightley (2009) found that on average 5% of amino-acid substitutions were adaptive if recent population bottlenecks were taken into account.

Another method to detect the presence of advantageous mutations is through investigating the underlying distribution of fitness effects among mutations. Using such a method Shaw et al. (2002) suggested that half of all mutations in *Arabidopsis thaliana* increased fitness (although see Keightley and Lynch (2003)). Even in fairly laboratory-adapted strains of *Saccharomyces cerevisiae*, Joseph and Hall (2004) estimated that around 6% of spontaneous mutations were beneficial (see also Hall and Joseph, 2010).

The strength of selection acting on beneficial alleles is also subject to much debate and is expected to depend on the nature of past environmental changes, both biotic and abiotic (Elena and Lenski, 2003). On the lower end, Jensen et al. (2008) estimated that advantageous mutants have had a mean selection coefficient of $s_a \approx 10^{-4}$ in *Drosophila*. On the upper end, very large selection coefficients have been detected in experimental evolution studies with bacteria and viruses, with an average $s_a \approx 2$ found in *Pseudomonas fluorescens* exposed to a novel carbon source (Barrett et al., 2006) and s_a ranging between 6 and 14 in the bacteriophage ϕX174 subjected to heat stress (Bull et al., 2000).

Although there is increasing evidence for the frequent spread of advantageous alleles, it is an inescapable fact that most spontaneous mutations that affect fitness are deleterious (Crow, 1970) and are maintained in populations at a low frequency by recurrent mutation (Wright, 1931). These mutation rates can be substantial; for example, the per-generation genomic deleterious mutation rate U_d in *Drosophila* has been estimated at 1.2 (Haag-Liautard et al., 2007; Keightley et al., 2009), with estimated rates of U_d of around 4.2 in hominids (Eöry et al., 2010). Deleterious mutation rates are lower in microbes, however. In nonmutator strains of yeast, Hill and Otto (2007) estimated $U_d = 0.013$ for mutations acting on sporulation ability and $U_d = 0.0003$ for those affecting growth rate.

If selection acts against deleterious mutations with a coefficient of s_d, then we would expect a total of $\sim U_d/s_d$ mutations to segregate within a population at mutation–selection balance (ignoring genetic associations among them). Even when U_d is less than one, the expected number of deleterious mutations carried by an individual may be much greater than one. Consequently, newly arisen advantageous alleles may occur within chromosomes also bearing deleterious alleles nearby. In the next section, we develop a model that describes the fate of a deleterious mutation that occurs in the genetic background of a novel beneficial allele. We later return to estimates of mutation rates and selection coefficients to assess how likely it is that deleterious alleles hitchhike to fixation, and how this depends on the mode of reproduction and the effective rate of recombination within a species.

Semi-Deterministic Model

We first present a semi-deterministic calculation of the fixation probability of a haplotype carrying both an advantageous and a deleterious allele using classic population genetics. In the next section, we build a stochastic diffusion model of the appearance and spread of this haplotype, but the calculations presented in this section help to develop an understanding of the key forces at work and so are a natural first step in investigating this problem.

We consider a finite population of N haploid chromosomes with discrete generations, using a standard Wright–Fisher model (Fisher, 1930; Wright, 1931). We are interested in the dynamics of a newly arisen beneficial allele at a locus A. The genome in which A first arises may carry one or more deleterious alleles. Deleterious alleles that are only loosely linked to locus A are unlikely to rise

Table 1. Table of haplotypes.

Haplotype	Fitness, w	Locus 1	Locus 2
$A_0\ B_0$	1	Wild type	Wild type
$A_1\ B_0$	$1 + s_a$	Beneficial	Wild type
$A_0\ B_1$	$1 - s_d$	Wild type	Deleterious
$A_1\ B_1$	$1 + s_a - s_d$	Beneficial	Deleterious

substantially in frequency and are ignored. We focus only on the single most closely linked of these deleterious mutations and call this second locus B, with recombination between A and B occurring at rate r. At locus A, the advantageous allele A_1 has a selective advantage s_a over the wild-type allele A_0. At locus B, the deleterious allele B_1 is selected against with selection coefficient s_d, relative to the wild-type allele B_0. We assume $s_a > s_d$, so that the advantageous-deleterious haplotype has a net beneficial effect, $s_{net} = s_a - s_d$. For clarity of presentation, we assume additive selection, but all of our analytical results continue to apply if s_d is replaced by $s_a - s_{net}$, wherever it occurs.

For each haplotype, we write a 0 subscript if the wild-type allele is present at the locus and a 1 subscript otherwise, in the order AB. All possible haplotypes, along with their fitness, are given in Table 1. In particular, the advantageous-deleterious haplotype is denoted A_1B_1, and when this haplotype first appears, the remainder of the population is either A_0B_0 (wild type) or A_0B_1 (bearing the deleterious allele). The latter haplotype (A_0B_1) is assumed to be rare and is ignored in the following analysis to simplify the calculations; simulations described in a later section indicate that this assumption introduces little bias. We also assume that no further mutation occurs at either of the loci during the course of the sweep, although the model can be modified to take this into account.

Let $p(t)$ denote the frequency of the A_1B_1 haplotype, where t is the number of generations since the beneficial allele arose and p_0 is its initial frequency (generally $1/N$). When the A_1B_1 haplotype first arises, it becomes established within the population with a probability u that is approximately twice the net selection coefficient, $2s_{net}$ (Haldane, 1927). It is further assumed that $s_{net} \ll 1$ and that the population size is large (see next section for results that apply in smaller populations).

In the following derivation, we only consider those A_1 alleles that survive stochastic loss while rare. Once established, the frequency of A_1B_1 can be modeled by the standard deterministic equation for haploid selection (Haldane, 1924):

$$p(t) = \frac{p_0(1 + s_{net})^t}{p_0(1 + s_{net})^t + 1 - p_0}. \quad (1)$$

Among those alleles that succeed in fixing, the trajectory of the A_1B_1 haplotype is slightly faster, on average, than given by equation (1) (Maynard Smith and Haigh, 1974; Barton, 1994). This initial acceleration is taken into account in the diffusion model developed below; it turns out to have little effect, however, because rare recombination events that break apart the A_1B_1 haplotype are most likely to occur when the A_1B_1 haplotype is intermediate in frequency and not when it initially occurs.

Our goal is to calculate the probability, P, that the A_1B_1 haplotype is not broken apart by recombination before the advantageous A_1 allele fixes within the population. If such a recombination event has not yet occurred, there are approximately $p(t)$ of the A_1B_1 haplotypes and $1 - p(t)$ of the A_0B_0 haplotypes (ignoring the rare A_0B_1 individuals), so that matings between these two haplotypes occur at frequency $2p(t)(1 - p(t))$. Among the offspring of these matings, r will be recombinant, half of which will carry the most fit A_1B_0 haplotype and half of which will carry the least-fit A_0B_1 haplotype. Even once produced, the most fit recombinant may fail to establish itself within the population due to chance loss while rare. In Appendix A, we use branching processes to show that the probability that a single new A_1B_0 haplotype establishes within the population if it appears at time t equals

$$\Pi(t) = \frac{2s_a s_d}{s_a p(t) + s_d(1 - p(t))} + O(s^2). \quad (2)$$

The derivation of equation (2) accounts for the fact that the A_1B_0 haplotype has fitness $1 + s_a$ relative to the population mean fitness $1 + p(t)(s_a - s_d)$, which is changing over time according to equation (1). As expected, if the A_1B_0 recombinant haplotype arises while $p(t) \approx 0$, the recombinant lineage will establish with probability nearly equal to $2s_a$, the fixation probability of an advantageous A_1 allele in an otherwise wild-type population. Also as expected, if the A_1B_0 recombinant haplotype arises while $p(t) \approx 1$, the recombinant lineage will establish with probability nearly equal to $2s_d$, the fixation probability of a haplotype that has shed the deleterious allele B_1 in a population that otherwise carries both A_1 and B_1. We call A_1B_0 haplotypes that succeed in establishing while rare "successful recombinants."

Altogether, $\kappa(t) = rp(t)(1 - p(t))\Pi(t)$ is the probability that an A_1B_0 recombinant haplotype appears at time t and goes on to establish within the population. Note however that this calculation does not specify whether the A_1 or B_0 allele will fix first; in many cases, if a recombinant appears and fixes with probability $\Pi(t)$, the actual fixation of the A_1B_0 haplotype would occur after A_1 has reached fixation.

To calculate the overall probability, P, that the A_1B_1 haplotype is never broken apart by recombination, we must calculate the probability that in every generation, t, none of the N offspring are successful recombinants. Assuming weak selection such that both $\Pi(t)$ and $\kappa(t)$ are small, the probability that a deleterious hitchhiker will be carried to fixation by the spread of a linked

beneficial allele is given by

$$\begin{aligned} P &= \prod_{t=0}^{\infty}(1-\kappa(t))^N \\ &\approx \prod_{t=0}^{\infty} exp[-N\kappa(t)] \\ &= exp\left[\sum_{t=0}^{\infty} -N\kappa(t)\right] \\ &\approx exp\left[\int_{t=0}^{\infty} -N\kappa(t)dt\right]. \end{aligned} \tag{3}$$

Overall, P gives the probability that a fitter recombinant never establishes, assuming that the A_1B_1 haplotype is not lost stochastically when it first appears. The probability that the A_1B_1 haplotype succeeds in establishing initially and fixing within the population is thus u ($= 2s_{net}$) times P. This equation is analogous to equation (16) in Yu and Etheridge (2010), who used a Moran model to estimate the fixation probability of two competing beneficial mutations, with recombination between the two loci.

Equation (3) can be solved by integrating over the allele frequency dynamics rather than over time and replacing the integral with

$$\int_{p=p_0}^{1} -\frac{N\ \kappa(p)}{dp/dt}\,dp. \tag{4}$$

In this haploid model with weak selection, $dp/dt = (s_a - s_d)\ p(1-p)$. Carrying out the integration, the probability that a fitter recombinant never establishes is given by

$$P \approx exp\left[-\frac{2\ N\ r\ s_a\ s_d\ \ln(s_a/s_d)}{(s_a - s_d)^2}\right],$$

where p_0 was assumed negligible relative to terms on the order of one. At this point, we can eliminate the population size from the result by measuring the net selection and recombination rates within the population, defined as $S_d = Ns_d$, $S_a = Ns_a$, $S_{net} = N(s_a - s_d)$, and $\rho = Nr$, yielding

$$P \approx \left(\frac{S_a}{S_d}\right)^{-\omega}, \tag{5}$$

where ω is the compound parameter defined by

$$\omega = 2\ \rho\ \frac{S_a\ S_d}{S_{net}^2}. \tag{6}$$

The hitchhiking process thus depends primarily on these scaled parameters and not separately on the population size and selection or recombination parameters. The above equations show that the probability of hitchhiking to fixation declines exponentially with the recombination rate between the loci and with the number of individuals within the population. The probability of hitchhiking is especially small when the strength of selection for the beneficial allele and against the deleterious allele is similar (S_{net} small), as this will cause the sweep of the A_1B_1 haplotype to take longer and allow for more recombination events.

To determine how small the recombination rate must be in order for hitchhiking to occur with a particular probability of interest, c, we set $P = c$ and solve for ρ:

$$\rho_{crit} = \frac{S_{net}}{S_d}\left[\frac{\ln\left(\frac{1}{c}\right)}{2\left(1+\frac{S_d}{S_{net}}\right)\ln\left(1+\frac{S_{net}}{S_d}\right)}\right]. \tag{7}$$

This gives us the recombination rate below which hitchhiking to fixation will occur with frequency greater than c, as a function only of the scaled selection coefficients S_d and S_{net}. At this point, we hold off discussing these results further until the next section, where we derive a stochastic solution.

Stochastic Model

The above analysis assumes that the population is very large, allowing us to combine stochastic results for the establishment of particular haplotypes while rare, with deterministic equations for the spread of these haplotypes. The above does not, however, take into account chance fluctuations in haplotype frequencies or the initial acceleration caused by considering only those trajectories where the beneficial allele becomes established (Maynard Smith and Haigh, 1974; Barton, 1994; Otto and Barton, 1997; Desai and Fisher, 2007). To account for these effects, we now derive a stochastic solution for this problem.

Again ignoring the rare deleterious-only lineage, we model the change in frequency, $p(t)$, of the A_1B_1 haplotype using a diffusion approximation. If a successful recombinant appears, however, the diffusion process is killed. As described by Karlin and Taylor (1981), the probability that the process is not ultimately killed, $P(p)$, given that A_1B_1 is currently at frequency p, satisfies

$$\frac{1}{2}V(p)\frac{d^2P(p)}{dp^2} + M(p)\frac{dP(p)}{dp} - K(p)P(p) = 0, \tag{8}$$

where $M(p)$ is the mean change in p over a time step measured in N generations; $V(p)$ is the variance in change of p; and $K(p)$ is the killing function, which denotes the probability of the process being "killed" while the A_1B_1 haplotype is at frequency p. In this model, killing occurs if recombination forms a fitter haplotype (i.e., A_1B_0) that succeeds in establishing within the population. To solve equation (8), we use the boundary conditions $P(0) = P(1) = 1$; that is, the system cannot be killed if the A_1B_1 or A_0B_0 haplotype is fixed. Further descriptions of similar diffusion models with killing are available in Karlin et al. (1967) and section 15.10 of Karlin and Taylor (1981); in particular, a related model is described where the diffusion process is killed whenever any recombinant is formed (A_1B_0 or A_0B_1), regardless of whether the recombinant succeeds in establishing.

As with standard diffusion models investigating an allele under weak directional selection in a haploid population (Kimura, 1970; Ewens, 2004), we obtain the values $M(p) = S_{net}\,p(1-p)$ and $V(p) = p(1-p)$, where $S_{net} = N(s_a - s_d)$ (see section 2 of Appendix S3). The killing term is obtained by taking the probability that the process is killed in a particular generation, $1-(1-\kappa)^N \approx N\kappa = N\,r\,p(1-p)\,\Pi$, and scaling in such a way that the killing term remains finite over the time step of N generations, as $N \to \infty$ (Karlin and Taylor, 1981). By doing so, we obtain the killing function $K(p) = \rho\,p(1-p)\,\pi(p)$, where $\rho = Nr$ and $\pi(p)$ is the scaled version of the establishment probability of the A_1B_0 recombinant, Π (eq. 2)

$$\pi(p) = \frac{2\,S_d\,(S_{net} + S_d)}{p\,S_{net} + S_d}. \tag{9}$$

The diffusion approximation assumes that S_{net}, S_d, and ρ remain finite as $N \to \infty$.

Plugging these diffusion coefficients into equation (8) and dividing by $p(1-p)$, the probability that the process is not killed, $P(p)$, given the current frequency p satisfies

$$\frac{1}{2}\frac{d^2P(p)}{dp^2} + S_{net}\frac{dP(p)}{dp} - \rho\,\pi(p)\,P(p) = 0. \tag{10}$$

If the process is not killed, there are two potential outcomes: fixation of A_0B_0 or fixation of A_1B_1. If we wish to know the probability that a particular advantageous allele that succeeds in fixing carries along with it a deleterious allele, we must rederive the diffusion model conditional on A_1 establishing within the population. In Appendix B, we show that the conditional probability $P^*(p)$ that the process is not killed (i.e., the deleterious allele B_1 fixes) among those cases where A_1 sweeps to fixation satisfies:

$$\frac{1}{2}\frac{d^2P^*(p)}{dp^2} + S_{net}\,\frac{1+e^{-2pS_{net}}}{1-e^{-2pS_{net}}}\frac{dP^*(p)}{dp} - \rho\,\pi(p)\,P^*(p) = 0. \tag{11}$$

The differential equations (10) and (11) were solved in *Mathematica* 6.0 (Supporting information), yielding the somewhat cumbersome equations (B5) and equation (B6), respectively. These can be solved numerically for the probability that the process is not ultimately killed (i.e., the probability that a successful recombinant never appears).

$P^*(p_0)$ as given by (B6) is the main quantity of interest in this article. It describes the probability that an A_1 allele that fixes within a population carries along with it a linked deleterious allele B_1, given that the initial frequency of the A_1B_1 haplotype is p_0. Although equations (B5) and (B6) should be used in any numerical analysis, further insight is provided by approximating $P*(p_0)$ as an exponentially decreasing function of the recombination rate (as inferred in the semi-deterministic analysis). Assuming that selection is strong relative to drift ($S_d, S_{net} \gg 1$), that the frequency of the A_1B_1 haplotype when the A_1 allele first appears is negligibly small ($p_0 \ll 1$), and that recombination is not too frequent ($\rho \ll S_d, S_{net}$), we obtain:

$$P^*(p) \approx \left(e^{-1/S_d}\,\frac{S_a}{S_d}\right)^{-\omega} \tag{12}$$

(see details in section 3 of Appendix S3). Again, this can be used to calculate a critical value of recombination above which hitchhiking is unlikely to occur. Specifically, we solve equation (12) for the rate of recombination necessary for the deleterious B_1 allele to fix with probability c, given that the beneficial allele A_1 initially appears with B_1 and ultimately fixes

$$\rho_{crit} = \frac{S_{net}}{S_d}\left[\frac{\ln\left(\frac{1}{c}\right)}{2\left(1+\frac{S_d}{S_{net}}\right)\left(\ln\left(1+\frac{S_{net}}{S_d}\right) - 1/S_d\right)}\right]. \tag{13}$$

For example, when $c = 1/2$, the term in square brackets is approximately 1/4 as long as neither S_d nor S_{net} is too small (see the figure in section 3 of the Appendix S3). Thus, as a rough rule of thumb (using unscaled parameters), the recombination rate r must be less that 1/4 of $s_{net}/(Ns_d)$ for there to be at least a 50% chance that the deleterious allele hitchhikes to fixation.

Hitchhiking events are thus likely to occur over larger regions of the genome if the net selection coefficient acting on the A_1B_1 haplotype, s_{net}, is stronger because sweeps occur faster. Conversely, the stronger the disadvantage of the deleterious allele, s_d, the less likely a hitchhiker will fix because recombinant A_1B_0 haplotypes are so much more fit. Finally, the larger the population size, the less likely that a hitchhiker will fix, simply because there are more individual chances for recombination to occur while the population remains polymorphic.

These patterns are illustrated in Figure 1, which gives the probability that the deleterious B_1 allele hitchhikes to fixation given that the beneficial A_1 allele fixes, with darker shading corresponding to higher probabilities. These contour plots are based on the exact solution (B6) to the diffusion equation for $P^*(p)$. The thick dashed curves show the approximate equation (13) for the critical value of the recombination rate, ρ, below which we expect deleterious alleles to hitchhike to fixation more than c of the time ($c = 10\%$, 50%, or 90%) when they occur on the haplotype bearing a new beneficial allele; these curves accurately follow the appropriate contour lines as long as selection is not too weak (roughly, $S_{net}, S_d \geq 2$).

COMPARISON TO THE CASE OF A LINKED NEUTRAL ALLELE

The dynamics of neutral loci are likely to be affected by the spread nearby of a beneficial allele whenever r is approximately less than s_a (Maynard Smith and Haigh, 1974). This rule cannot be used to compare to equation (13) directly, however, because our criteria for being "affected" is now quite strict: the linked B_1 allele must fix due to the sweep. We thus briefly describe a corresponding

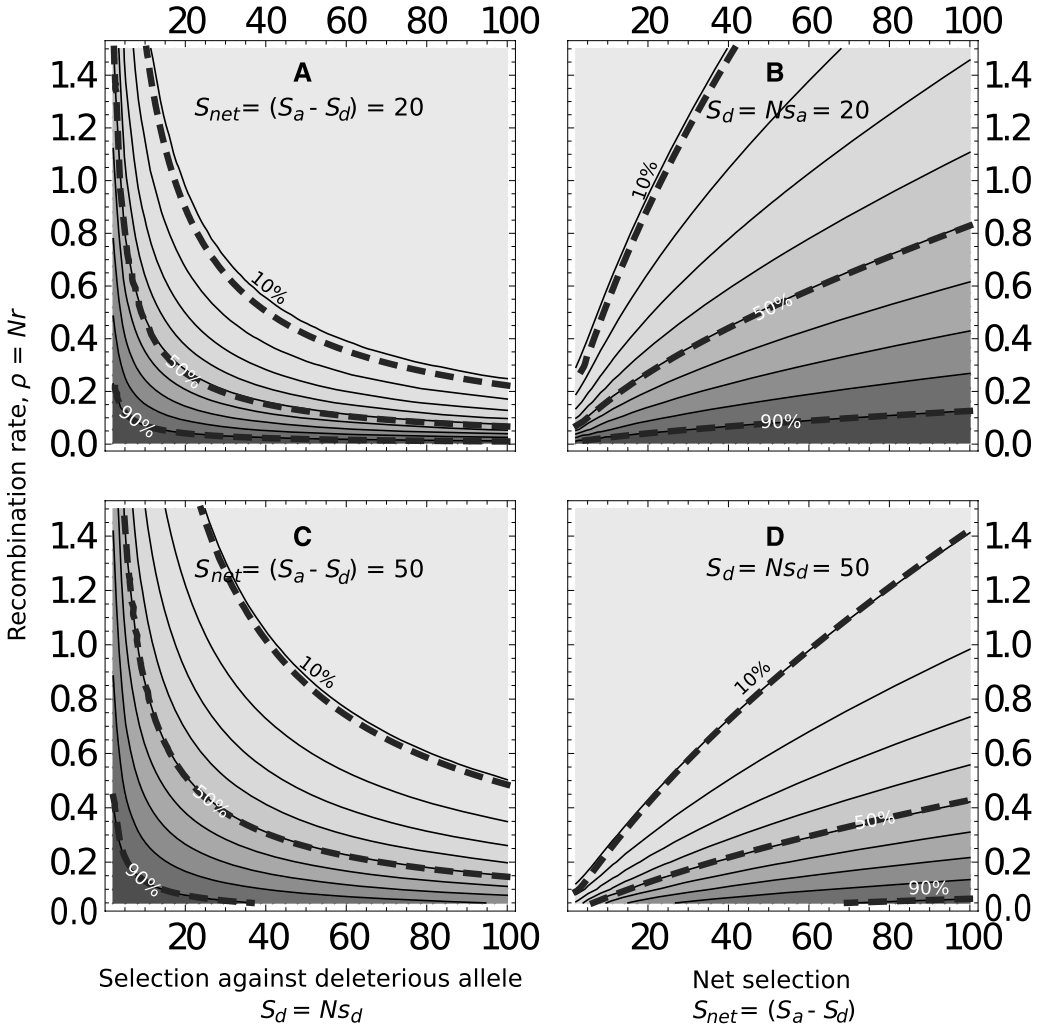


Figure 1. Contour plots of the fixation probability of the deleterious B_1 allele, given that the A_1B_1 haplotype appears initially at frequency of $1/N$ and that the A_1 allele is not lost stochastically (10% contour intervals based on equation B6). The graphs are shown for N = 10,000, although the results are not very sensitive to N, as long as the scaled parameters are held constant. In each case, ρ is plotted along the y-axis versus S_d along the x-axis (left panels) with S_{net} = 20 (top) or 50 (bottom) or versus S_{net} along the x-axis (right panels) with S_d = 20 (top) or 50 (bottom). The dashed curves show the predicted thresholds below which there is a greater than c = 10%, 50%, and 90% probability of hitchhiking, based on equation 13; in each case this threshold coincides closely with the appropriate contours.

model for the case when B is neutral (full details are provided in the section 4 of Appendix S3).

The diffusion equations remain essentially the same, except that the killing term must be revised now that the recombinant A_1B_0 haplotype is no more fit than the A_1B_1 haplotype that is spreading through the population. We assume that, whenever a recombinant A_1B_0 haplotype appears, the probability that this haplotype becomes the ancestor of the population at some distant future point in time is very nearly $1/(Np)$. This assumes that any individual carrying the A_1 allele alive at that time is equally likely to be the lucky one to ultimately fix and give rise to the entire descendant population. Using $1/(Np)$ instead of Π for the fixation probability of the recombinant A_1B_0 haplotype, we obtain the revised killing function, $K(p) = \rho\, p(1 - p)\, 1/p$, for use in the diffusion equation (8), assuming that allele A_1 fixes. The conditional probability of the process not being killed was then obtained using *Mathematica* 6.0.

Focusing on the conditional probability that the process reaches fixation on A_1 before being killed by the appearance of a successful recombinant, we again obtained an approximation assuming that selection is strong relative to drift

$$P^*(p_0) = (2\, e^{\gamma}\, S_{net})^{-\rho/S_{net}}, \tag{14}$$

where $\gamma = 0.577$ is Euler's constant. We have persisted in referring to the net selection on the A_1B_1 haplotype as S_{net} despite the fact that now $S_{net} = S_a$ for ease of comparison with the previous case.

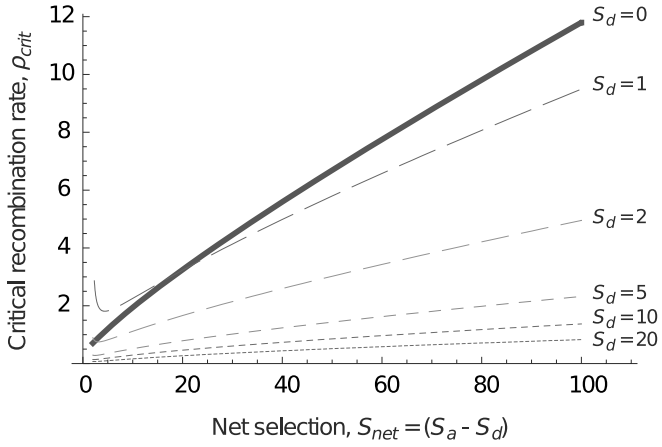


Figure 2. The critical value of the recombination rate, ρ_{crit}, below which there is a greater than $c = 50\%$ probability that the deleterious B_1 allele will hitchhike to fixation along with the advantageous allele, as a result of their initial association based on the approximate equations (13) and (15). In each case, ρ_{crit} is plotted along the *y*-axis versus S_{net} along the *x*-axis, for varying values of S_d. The case of a neutral linked allele at the *B* locus is given by the thick top curve. (The upturns in some of the curves near the origin as well as crossing of some of the curves are caused by inaccuracies in these approximations when selection is weak relative to drift.)

Again, solving this equation for the critical value of ρ below which hitchhiking to fixation occurs more than a proportion c of the time, we get

$$\rho_{crit}^{neutral} = S_{net}\left[\frac{\ln\left(\frac{1}{c}\right)}{\gamma + \ln(2\ S_{net})}\right]. \qquad (15)$$

For $c = 1/2$, the term in square brackets is approximately 1/4 when $S_{net} = 5$, and it continues to decline (but slowly) as S_{net} increases. Thus, as a rough rule of thumb, r must be less than $\approx$1/4 of s_{net} for there to be a 50% chance that a neutral allele hitchhikes to fixation. Again, such hitchhiking events are likely to occur over larger regions of the genome when the sweeps are faster (s_{net} large). The key difference, however, from the case with a deleterious hitchhiker is the absence of Ns_d in the denominator of this rule, which makes it easier to satisfy than the case of a deleterious hitchhiker (assuming selection is strong relative to drift). Figure 2 shows just how much more likely it is for alleles at locus B to hitchhike to fixation along with allele A when the B locus is neutral (thick top curve) than when it is subject to selection against deleterious mutations (dashed curves).

The fact that neutral alleles are much more likely to hitchhike to fixation than linked deleterious alleles has another important implication. Namely, the presence of a linked deleterious allele increases the chance that surrounding genetic variation will be rescued by recombination. Had there been no linked sites under selection, we would expect a region surrounding a sweep to be entirely fixed when $\rho < \rho_{crit}^{neutral}$ in the majority of cases (eq. 15). If a beneficial allele first occurs on a chromosome containing a deleterious allele, however, this region is greatly reduced to $\rho < \rho_{crit}$ (eq. 13), as illustrated in Figure 2. Consequently, linkage to sites carrying deleterious alleles reduces the impact of selective sweeps, making it less likely that surrounding genetic variation will be lost.

Turning this argument around, a recently fixed beneficial allele might have been strongly selected but appear to have been weakly selected based on the amount of genetic variation remaining in the region. This is because recombinants were favored that untied the beneficial allele from the deleterious genetic baggage with which it arose. Furthermore, we would expect that genetic variation should more often be rescued by the appearance of more fit recombinants on the side of a selective sweep that bears a higher density of other sites under selection. In Supporting information, we simulate a three-locus model with one locus subject to advantageous mutation, one locus being a neutral marker, and one locus subject to recurrent deleterious mutation, with the beneficial mutant placed on a randomly selected genetic background. As confirmed in Figure S1 the sweep of neutral diversity is less severe in cases where selection acts on the locus subject to deleterious mutations.

TWO COMPETING BENEFICIAL MUTATIONS

The above analyses can also be used to solve a related problem of beneficial mutations competing for fixation in the presence of recombination, as considered by Yu and Etheridge (2010). If a beneficial allele is rising in frequency when a second beneficial allele appears at a linked site, then it is possible for the first beneficial allele to be lost if the second allele is more strongly favored if it appears with the wildtype allele at the first locus, and if a recombinant that brings together both alleles onto the same haplotype fails to establish in time.

Although technically there are three chromosome types to be considered before the recombinant appears (00, 10, and the new 01, where the "1" now indicates a beneficial mutation at the first and second sites), we can approximate this scenario as did Yu and Etheridge (2010) by assuming that the 00 wild type is rapidly eliminated, so that the frequencies of 01 and 10 sum roughly to one. This approximation performs surprisingly well for this problem because rare recombination events do not occur until the 10 and 01 haplotypes are both common.

Equation (1) then describes the spread of the more fit 01 haplotype, whose frequency is $\approx p(t)$ (frequency of 10 $\approx 1 - p(t)$), with s_{net} equal to the difference in fitness between 10 and 01 individuals. Equation (2) describes the fixation probability of a recombinant double mutant, with s_a and s_d giving the selective advantage of the double mutant when it appears in a population

predominantly composed of 10 and 01 individuals, respectively. All of the subsequent results described above then follow. Figure S4 shows that equations (5) and (12) provide an excellent estimate of the probability that recombination successfully rescues both beneficial mutations. Although similar in spirit to the work of Yu and Etheridge (2010), our analyses have the advantage of providing closed-form solutions that appear to accurately capture the stochastic nature of recombination rescuing combinations of beneficial alleles at two selected loci.

Two-locus simulations

To investigate the accuracy of the above results, we compare both the semi-determinstic and stochastic models to Monte Carlo simulations. Simulations start with a population of N haploid chromosomes, each consisting of two linked loci. Fitness is assumed to be additive.

An initial proportion $p_0 = 1/N$ of the population is assigned the advantageous-deleterious A_1B_1 haplotype. The rest of the population bears the A_0B_0 haplotype. It is assumed that the A_0B_1 haplotype is present at a negligibly small frequency, and while it is not considered in the initial population it is tracked if it appears by recombination.

A new generation is formed by selecting two parents with probability proportional to their fitness. Recombination between the two parental loci then occurs with Poisson probability r. This is repeated until N new offspring are created. A new generation is created in this way until the A_1B_1 genotype is either fixed or is lost from the population. This entire process is repeated 20,000 times to build up an overall probability of fixation along with 95% confidence intervals. We focus attention on the processes where the advantageous allele fixes.

Results are plotted in Figure 3. Simulation data match up very well to all three solutions for the probability of hitchhiking $P^*(p)$: semi-deterministic equation (5), diffusion equation (B6), and the approximation to the diffusion equation (12). All three solutions offer similar results when we changed the population size, as long as ρ, S_d, S_a are held constant. Differences between the solutions only become apparent when selection becomes weak. Stochastic effects then play more of a role, especially where the A_1B_1 haplotype is oversampled and rises to fixation faster than expected, so that the diffusion with killing (B6) provides a slightly more accurate solution. Additional figures presented in section 3 of the Appendix S3 show that the analytical solutions perform less well as selection strengthens in very small populations (e.g., $s_d = 0.1$ with N = 100 or 1000); in these cases, the diffusion approximation assuming weak selection breaks down and the fixation probability of the deleterious allele is underestimated.

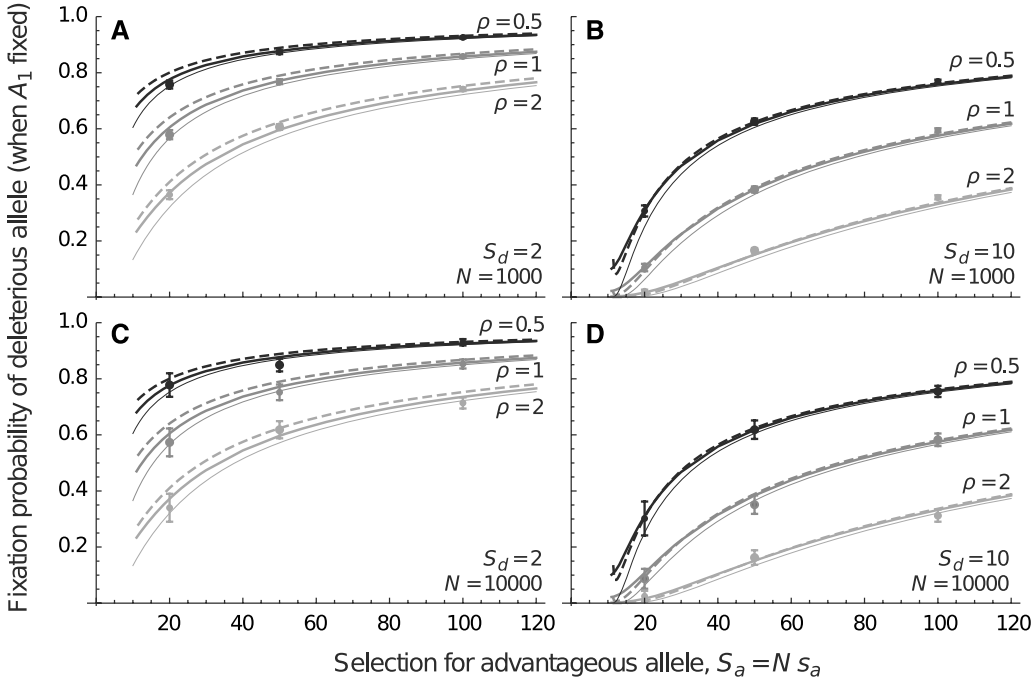


Figure 3. Fixation probability of the deleterious B_1 allele, given that the A_1B_1 haplotype appears initially at frequency of $1/N$ and that the A_1 allele is not lost when rare, for different recombination rates $\rho = Nr$. Plots compare the solution to the semi-deterministic model (5) (thin solid), the full solution to the diffusion (B6) (thick solid), the approximation to the diffusion (12) (thick dashed), and simulation results based on the Wright–Fisher model (points). Bars indicate 95% confidence intervals here and throughout. Parameters are $N = 1000$ (A and B) and $N = 10,000$ (C and D), with $S_d = 2$ (A and C) and $S_d = 10$ (B and D).

Multilocus simulations

Although the above two-locus models offer tractable results, novel advantageous alleles may arise in genomes with multiple mutant alleles. Therefore, we switch to using multilocus computer simulations to investigate the mutation load generated by the rise to fixation of an advantageous allele, given that such mutations arise at rate U in a genome with total map length R, where each new deleterious mutations is assigned a random position between 0 and R. The methods used for these simulations are based on Hartfield et al. (2010) and detailed in Supporting information.

We then determined the mean number of deleterious alleles that fix along with each beneficial mutation, assuming multiplicative selection. Simulations with different S_a values are compared to the control case, $S_a = 0$, in Figures 4 and S2. These results corroborate the two-locus model; the mean number of deleterious mutants that fix declines with the rate of recombination and rises with the strength of selection on the advantageous mutant, S_a. The mean number of fixed deleterious alleles also stays approximately the same as N increases, if the compound parameters S_a, S_d, NR, and NU are held constant.

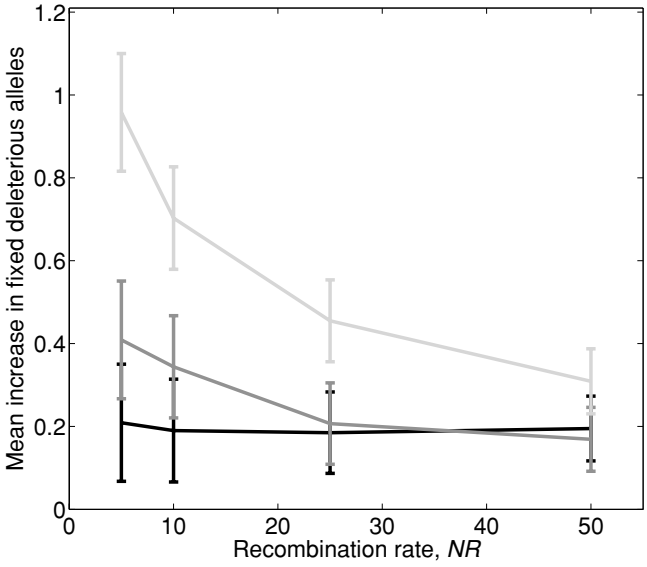
Increasing the recombination rate also raises the fixation probability of the advantageous mutant (Fig. S3), which is a well-known result (Peck, 1994; Barton, 1995). Thus recombination is doubly advantageous, as it reduces the number of deleterious alleles that fix in a population following a selective sweep and it increases the likelihood that such an advantageous mutant can establish when rare. This is the likeliest cause of strong selection acting on a modifier for increased recombination in the presence of advantageous and deleterious mutations (Hartfield et al., 2010).

APPLYING RESULTS TO HUMAN GENETIC DATA

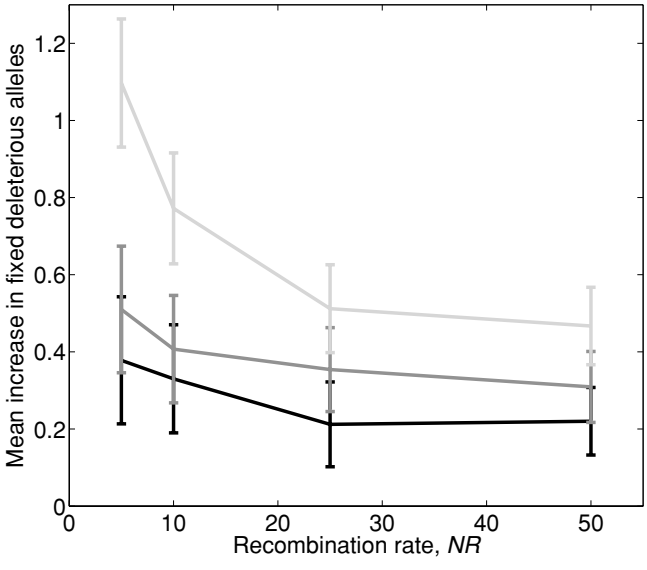
How likely is deleterious hitchhiking to occur in nature? To answer this, we use human data as an example. Deleterious mutants are maintained at a mutation–selection balance frequency of $q = \mu/s_d$ (Wright, 1931), where s_d measures selection against the deleterious allele in heterozygotes. Thus an estimate for the number of deleterious mutants segregating throughout a genome is U/s_d, for U the diploid per-genome deleterious mutation rate, which has been recently estimated as $U = 4.2$ (Eöry et al., 2010).

U measures deleterious mutations arising across the entire genome, with the majority appearing in noncoding regions (Eöry et al., 2010). Thus we assume all deleterious mutations have a fixed, weak value of s_d. This will slightly overestimate the number of deleterious mutants segregating, as we do not consider stronger deleterious mutations that can arise in coding regions (Eyre-Walker et al., 2006; Boyko et al., 2008).

A deleterious allele must have $N_e s_d \geq 1$ in order for selection to overcome the effects of genetic drift (Kimura, 1983). Therefore, assuming deleterious alleles are very weakly selected



(a)



(b)

Figure 4. The increase in the number of deleterious alleles that fix genome-wide for a given S_a, subtracting off the number that fix in the $S_a = 0$ case, as a function of the total map length NR (see Fig. S2 for the raw data). Only cases where the advantageous allele has fixed are considered. $S_a = 20$ (black line), 40 (dark gray line), or 80 (light gray line). $S_d = 10$, $NU = 50$, and (a) $N = 500$ or (b) $N = 1000$.

($N_e s_d = 1$, with human $N_e = 10{,}000$; Jorde et al. 1998), we expect $U/s_d = 4.2/0.0001 = 42{,}000$ such deleterious alleles segregating at any time, roughly half of which lie in each haploid set of 3 Gb in the human genome. Including the site of the beneficial mutation, the average distance between two selected sites is thus 142.9 kb. Assuming that selected sites are randomly distributed across

the genome (i.e., ignoring clustering), this distance would be approximately exponentially distributed. In this case, the closest of the deleterious alleles lying to either side of the beneficial allele would also be exponentially distributed with mean 71.4 kb. As a rough guide, the average recombination rate is 1 cM/Mb in a human genome (Broman et al., 1998), thus the closest deleterious allele lies, on average, at a distance of $N_e r = 7.14$. The fixation probability of the deleterious allele with the advantageous mutant would then be 18.8% for $N_e s_a = 5$, 37.1% for $N_e s_a = 25$, and 62.1% for $N_e s_a = 100$, obtained by integrating the hitchhiking probability (B6) over an exponentially distributed distance with mean $N_e r = 7.14$. These calculations are explained in more detail in section 5 of the Appendix S3. If we assumed $N_e s_d = 10$, then by following a similar logic we calculate that the mean distance to the nearest deleterious allele is $N_e r = 71.4$, and the estimated fixation probability of a deleterious allele is 0.8% for $N_e s_a = 25$ and 2.5% for $N_e s_a = 100$.

Overall, these calculations suggest that in humans, deleterious mutants will hitchhike at appreciable frequencies only if they are very weakly selected ($N_e s_d < 10$). However, this is only an initial calculation that deserves to be revised to take into account fine-scale recombination rates (McVean et al., 2004) and clustering of mutations around coding regions. For now we note that if clustering causes the average recombination distance to a deleterious allele to drop tenfold, then the hitchhiking probabilities calculated above increase substantially, rising for $N_e s_d = 1$ to 68%, 85%, 94% with $N_e s_a = 5$, 25, 100, respectively, and for $N_e s_d = 10$ to 7%, 20% with $N_e s_a = 25$ and 100.

Discussion

As long as genetic variance in fitness is present within a population, new beneficial alleles can arise in genomes that, by chance, carry deleterious alleles at linked sites. Consequently, if they remain associated, deleterious alleles can hitchhike to fixation as an advantageous allele sweeps through the population. Even if recombination occurs between the two loci, there can still be a good chance of both alleles fixing, if either the recombinant fails to appear in time or is lost by chance when it does appear. Williamson et al. (2007) found possible evidence of such hitchhiking causing the high prevalence of the hereditary hemochromatosis mutation C282Y, due to a selective sweep occurring 150 kb away from the HFE gene where the deleterious C282Y allele is located.

To our knowledge, this article represents the first theoretical study on how recombination affects the hitchhiking to fixation of deleterious alleles. Using both a semi-deterministic and a diffusion approach, we show that in regions of low recombination there is a high probability that a deleterious mutant would be swept to fixation if linked to an advantageous mutant (Fig. 1). This probability approaches one as the deleterious effect s_d tends towards zero and the overall advantage of the A_1B_1 haplotype s_{net} is larger. Outside this parameter range, we find that hitchhiking is likely (greater then 50% chance) if $r \lesssim s_{net}/(4Ns_d)$ (more precisely, equation 13). A promising empirical approach would be to investigate areas around the genome that show high d_N/d_S values. Such regions are assumed to be subject to recurrent sweeps (Nielsen, 2005). If deleterious alleles do hitchhike, then around these sites there should be signs of increased load, such as increased indel frequency, or lower frequency of optimal codon usage. Such a negative relationship between *dN* and optimal codon usage was found in *Drosophila* by Betancourt and Presgraves (2002).

Furthermore, we determined that the hitchhiking of tightly linked deleterious alleles reduces the region in which the sweep is likely to fix surrounding sites (compare eq. 15 to eq. 13). This is important as it implies that deleterious hitchhiking can alter experimental estimates of the strength of such sweeps. A potential example of these effects was reported by Clegg et al. (1980), who found that linkage disequilibrium in *D. melanogaster* broke down more quickly than expected (geometric decay at a ratio $1 - r$), based on the surrounding markers being neutral and on measured recombination rates between the selected and neutral markers. This observation could be explained by recombination untangling advantageous alleles from deleterious backgrounds (see also Fig. S1). Further work is warranted to explore the impact of neighboring selected sites on patterns of neutral sequence variability in a fully multilocus framework. In particular, a full treatment requires an exploration not only of the primary effects of a selective sweep at a focal site, but also of how hitchhiking of deleterious alleles can cause secondary sweeps as wild-type alleles reestablish themselves at surrounding sites.

Our work also sheds light on the results found by Hartfield et al. (2010), who showed that a modifier gene for increased recombination is more likely to fix in a population that is subject to both deleterious and advantageous mutation, compared to the deleterious-only mutation case (Keightley and Otto, 2006). The increased selection acting on a recombination modifier when both deleterious and advantageous mutants are present together, compared to when just deleterious or just advantageous mutations are present, suggests that uncoupling advantageous mutants from deleterious backgrounds provides a substantial amount of selection on a recombination modifier (Peck, 1994; Hartfield et al., 2010).

Our preliminary calculations suggest that in obligately sexual species with long genetic map lengths (such as the human genome), recombination is frequent enough to prevent all but weakly deleterious mutants from hitchhiking with advantageous mutants. Our calculations assumed, however, that mutations affecting fitness arise at equal rates throughout the genome, which ignores the clustering of fitness-impacting sites near genic regions. If recombination rates between selected sites are low, either

because of this clustering or because of cold spots in recombination, the probability that deleterious alleles hitchhike to fixation rises substantially. Similarly, in species that frequently inbreed (e.g., selfing) or reproduce asexually, the effective amount of recombination may be much lower, substantially increasing the probability of deleterious alleles hitchhiking to fixation. In asexuals with no recombination, the subsequent mutation accumulation can be extremely detrimental (Hadany and Feldman, 2005).

In conclusion, sex and recombination both enhance the probability of beneficial alleles establishing and hinder the fixation of deleterious alleles within a lineage. If this can be shown empirically to be a potent selective force on recombination rates, then this would provide key insight into why sex and recombination are prevalent, which remains an open question in evolutionary genetics (Otto, 2009).

ACKNOWLEDGMENTS

We would like to thank P. Keightley for his support and advice on using human genetic data. We also thank P. Keightley, N. Barton, J. Hermisson, and two anonymous referees for comments on the manuscript. MH is funded by a Biotechnology and Biological Sciences Research Council studentship; SO is funded by the Natural Sciences and Engineering Research Council of Canada.

LITERATURE CITED

Abramowitz, M., and I. Stegun. 1970. Handbook of mathematical functions. Dover Publications, Inc., New York.

Andolfatto, P. 2005. Adaptive evolution of non-coding DNA in *Drosophila*. Nature 437:1149–1152.

———. 2007. Hitchhiking effects of recurrent beneficial amino acid substitutions in the *Drosophila melanogaster* genome. Genome Res. 17:1755–1762.

Bachtrog, D., and I. Gordo. 2004. Adaptive evolution of asexual populations under Muller's ratchet. Evolution 58:1403–1413.

Barrett, R. D. H., R. Craig MacLean, and G. Bell. 2006. Mutations of intermediate effect are responsible for adaptation in evolving *Pseudomonas fluorescens* populations. Biol. Lett. 2:236–238.

Barton, N. H. 1994. The reduction in fixation probability caused by substitutions at linked loci. Genet. Res. 64:199–208.

———. 1995. Linkage and the limits to natural selection. Genetics 140:821–841.

Barton, N. H., and S. P. Otto. 2005. Evolution of recombination due to random drift. Genetics 169:2353–2370.

Betancourt, A. J., and D. C. Presgraves. 2002. Linkage limits the power of natural selection in *Drosophila*. Proc. Natl. Acad. Sci. USA 99:13616–13620.

Bierne, N., and A. Eyre-Walker. 2004. The genomic rate of adaptive amino acid substitution in *Drosophila*. Mol. Biol. Evol. 21:1350–1360.

Boyko, A. R., S. H. Williamson, A. R. Indap, J. D. Degenhardt, R. D. Hernandez, K. E. Lohmueller, M. D. Adams, S. Schmidt, J. J. Sninsky, S. R. Sunyaev, et al. 2008. Assessing the evolutionary impact of amino acid mutations in the human genome. PLoS Genet. 4:e1000083.

Broman, K. W., J. C. Murray, V. C. Sheffield, R. L. White, and J. L. Weber. 1998. Comprehensive human genetic maps: individual and sex-specific variation in recombination. Am. J. Hum. Genet. 63:861–869.

Bull, J. J., M. R. Badgett, and H. A. Wichman. 2000. Big-benefit mutations in a bacteriophage inhibited with heat. Mol. Biol. Evol. 17:942–950.

Charlesworth, B., M. T. Morgan, and D. Charlesworth. 1993. The effect of deleterious mutations on neutral molecular variation. Genetics 134:1289–1303.

Clegg, M. T., J. F. Kidwell, and C. R. Horch. 1980. Dynamics of correlated genetic systems. *V. Rates of decay of linkage disequilibria in experimental populations of Drosophila melanogaster. Genetics* 94:217–234.

Crow, J. F. 1970. Genetic loads and the cost of natural selection. Pp. 128—177, *in* K. I. Kojima, ed. Mathematical topics in population genetics, *Biomathematics*, vol. 1. Springer-Verlag, Berlin.

Desai, M. M., and D. S. Fisher. 2007. Beneficial mutation-selection balance and the effect of linkage on positive selection. Genetics 176:1759–1798.

Elena, S. F., and R. E. Lenski. 2003. Evolution experiments with microorganisms: the dynamics and genetic bases of adaptation. Nat. Rev. Genet. 4:457–469.

Eöry, L., D. L. Halligan, and P. D. Keightley. 2010. Distributions of selectively constrained sites and deleterious mutation rates in the hominid and murid genomes. Mol. Biol. Evol. 27:177–192.

Ewens, W. J. 2004. Mathematical population genetics: 1. Theoretical introduction, *Interdisciplinary applied mathematics*, Vol. 27. 2nd ed. Springer, New York.

Eyre-Walker, A. 2006. The genomic rate of adaptive evolution. Trends Ecol. Evol. 21:569–575.

Eyre-Walker, A., and P. D. Keightley. 2009. Estimating the rate of adaptive molecular evolution in the presence of slightly deleterious mutations and population size change. Mol. Biol. Evol. 26:2097–2108.

Eyre-Walker, A., M. Woolfit, and T. Phelps. 2006. The distribution of fitness effects of new deleterious amino acid mutations in humans. Genetics 173:891–900.

Fisher, R. A. 1930. The genetical theory of natural selection. The Clarendon Press, Oxford.

Haag-Liautard, C., M. Dorris, X. Maside, S. Macaskill, D. L. Halligan, B. Charlesworth, and P. D. Keightley. 2007. Direct estimation of per nucleotide and genomic deleterious mutation rates in *Drosophila*. Nature 445:82–85.

Hadany, L., and M. W. Feldman. 2005. Evolutionary traction: the cost of adaptation and the evolution of sex. J. Evol. Biol. 18:309–314.

Haldane, J. 1924. A mathematical theory of natural and artificial selection, part I. Trans. Cambridge Philos. Soc. 23:19–41.

Haldane, J. B. S. 1927. A mathematical theory of natural and artificial selection, part V: Selection and mutation. Math. Proc. Cambridge Philos. Soc. 23:838–844.

Hall, D. W., and S. B. Joseph. 2010. A high frequency of beneficial mutations across multiple fitness components in *Saccharomyces cerevisiae*. Genetics 185:1397–1409.

Halligan, D. L., F. Oliver, A. Eyre-Walker, B. Harr, and P. D. Keightley. 2010. Evidence for pervasive adaptive protein evolution in wild mice. PLoS Genet. 6:e1000825.

Hartfield, M., S. P. Otto, and P. D. Keightley. 2010. The role of advantageous mutations in enhancing the evolution of a recombination modifier. Genetics 184:1153–1164.

Hill, J. A., and S. P. Otto. 2007. The role of pleiotropy in the maintenance of sex in yeast. Genetics 175:1419–1427.

Hill, W. G., and A. Robertson. 1966. The effect of linkage on limits to artificial selection. Genet. Res. 8:269–294.

Hudson, R. R., M. Kreitman, and M. Aguade. 1987. A test of neutral molecular evolution based on nucleotide data. Genetics 116:153–159.

Jensen, J. D., K. R. Thornton, and P. Andolfatto. 2008. An approximate Bayesian estimator suggests strong, recurrent selective sweeps in *Drosophila*. PLoS Genet. 4:e1000198.

Johnson, T., and N. H. Barton. 2002. The effect of deleterious alleles on adaptation in asexual populations. Genetics 162:395–411.

Jorde, L. B., M. Bamshad, and A. R. Rogers. 1998. Using mitochondrial and nuclear DNA markers to reconstruct human evolution. BioEssays 20:126–136.

Joseph, S. B., and D. W. Hall. 2004. Spontaneous mutations in diploid *Saccharomyces cerevisiae*: more beneficial than expected. Genetics 168:1817–1825.

Karlin, S., J. L. McGregor, and W. Bodmer. 1967. The rate of production of recombinants between linked genes in finite populations. Proc. Fifth Berkeley Symp. on Math. Stat. Prob. 4:403–414.

Karlin, S., and H. M. Taylor. 1981. A second course in stochastic processes. Academic Press, New York.

Keightley, P. D., and M. Lynch. 2003. Towards a realistic model of mutations affecting fitness. Evolution 57:683–685.

Keightley, P. D., and S. P. Otto. 2006. Interference among deleterious mutations favours sex and recombination in finite populations. Nature 443:89–92.

Keightley, P. D., U. Trivedi, M. Thomson, F. Oliver, S. Kumar, and M. L. Blaxter. 2009. Analysis of the genome sequences of three *Drosophila melanogaster* spontaneous mutation accumulation lines. Genome Res. 19:1195–1201.

Kimura, M. 1970. Stochastic processes in population genetics. Pp. 178–209, *in* K. -I. Kojima, ed. Mathematical Topics in Population Genetics, *Biomathematics*, vol. 1. Springer-Verlag; Heidelberg; New York, Berlin.

———. 1983. The neutral theory of molecular evolution. Cambridge Univ. Press, Cambridge.

Kimura, M., and T. Ohta. 1970. Probability of fixation of a mutant gene in a finite population when selective advantage decreases with time. Genetics 65:525–534.

Smith, J. Maynard, and J. Haigh. 1974. The hitch-hiking effect of a favourable gene. Genet. Res. 23:23–35.

McVean, G. A. T., S. R. Myers, S. Hunt, P. Deloukas, D. R. Bentley, and P. Donnelly. 2004. The fine-scale structure of recombination rate variation in the human genome. Science 304:581–584.

Nielsen, R. 2005. Molecular signals of natural selection. Annu. Rev. Genet. 39:197–218.

Obbard, D. J., J. J. Welch, K.-W. Kim, and F. M. Jiggins. 2009. Quantifying adaptive evolution in the *Drosophila* immune system. PLoS Genet. 5:e1000698.

Otto, S. P. 2009. The evolutionary enigma of sex. Am. Nat. 174:S1–S14.

Otto, S. P., and N. H. Barton. 1997. The evolution of recombination: removing the limits to natural selection. Genetics 147:879–906.

Peck, J. R. 1994. A ruby in the rubbish: Beneficial mutations, deleterious mutations and the evolution of sex. Genetics 137:597–606.

Rice, W. R. 1999. Free content genetic polarization: unifying theories for the adaptive significance of recombination. J. Evol. Biol. 12:1047–1049.

Roze, D., and N. H. Barton. 2006. The Hill-Robertson effect and the evolution of recombination. Genetics 173:1793–1811.

Shapiro, J. A., W. Huang, C. Zhang, M. J. Hubisz, J. Lu, D. A. Turissini, S. Fang, H.-Y. Wang, R. R. Hudson, R. Nielsen, Z. et al. 2007. Adaptive genic evolution in the *Drosophila* genomes. Proc. Natl. Acad. Sci. USA 104:2271–2276.

Shaw, F. H., C. J. Geyer, and R. G. Shaw. 2002. A comprehensive model of mutations affecting fitness and interferences for *Arabidopsis thaliana*. Evolution 56:453–463.

Thomson, G. 1977. The effect of a selected locus on linked neutral loci. Genetics 85:753–788.

Williamson, S. H., M. J. Hubisz, A. G. Clark, B. A. Payseur, C. D. Bustamante, and R. Nielsen. 2007. Localizing recent adaptive evolution in the human genome. PLoS Genet. 3:e90.

Wright, S. 1931. Evolution in Mendelian populations. Genetics 16:97–159.

Yu, F., and A. M. Etheridge. 2008. Rate of adaptation of large populations. Pp. 3–27, *in* P. Pontarotti, ed. Evolutionary Biology from Concept to Application. Springer-Verlag, Berlin.

———. 2010. The fixation probability of two competing beneficial mutations. Theor. Popul. Biol. 78:36–45.

Yu, F., A. M. Etheridge, and C. Cuthbertson. 2010. Asymptotic behavior of the rate of adaptation. Ann. Appl. Probab. 20:978–1004.

Associate Editor: J. Hermisson

Appendix A

DERIVATION OF $\Pi(t)$, THE PROBABILITY OF ESTABLISHMENT OF A RECOMBINANT HAPLOTYPE

When the recombinant A_1B_0 haplotype is produced, it appears within a population that is already changing due to the spread of the A_1B_1 haplotype. Thus, we cannot calculate the probability of fixation of the recombinant A_1B_0 haplotype based solely on its fitness $1 + s_a$ relative to the current population mean $1 + p(t)(s_a - s_d)$. Rather, we must also account for future changes in the population mean fitness as the A_1B_1 haplotype rises in frequency. To do so, we develop a time-inhomogeneous branching process that explicitly follows the dynamics of $p(t)$ (given by eq. 1) that occur after the appearance of the recombinant A_1B_0 haplotype. A previous diffusion analysis by Kimura and Ohta (1970) also calculated the fixation probability for a favorable allele whose benefit declined over time, but the focus of their analysis was on a case where selection declines linearly over time, whereas here the selection coefficient favoring A_1B_0 declines according to a logistic function of time, given by $s(t) = s_a - p(t)(s_a - s_d)$.

Let $\Pi(t)$ be the fixation probability of the recombinant A_1B_0 haplotype at generation t, given that the current frequency of the A_1B_1 haplotype is $p(t)$. In a population of constant size, the average parent has one surviving offspring, but we assume that the A_1B_0 haplotype is more fit and so has an average of $1 + s(t)$ offspring. Using branching process logic (Haldane, 1927), the recombinant A_1B_0 haplotype will ultimately be lost (with probability $1 - \Pi(t)$) if and only if all j offspring inheriting the haplotype also fail to leave any descendants over the long run (with probability $(1 - \Pi(t+1))^j$). Assuming a Poisson distribution for the number of offspring j and summing over this distribution, we obtain a recursion for $\Pi(t)$:

$$1 - \Pi(t) = \sum_{j=0}^{\infty} e^{-(1+s(t))} \frac{(1+s(t))^j}{j!} (1 - \Pi(t+1))^j \qquad \text{(A1)}$$
$$= exp[-(1 + s(t))\ \Pi(t+1)].$$

Solving for $\Pi(t+1)$ and subtracting $\Pi(t)$, we obtain the change in fixation probability over time, which we assume is slow enough that it can be well approximated by the differential

equation

$$\frac{d\Pi}{dt} = -\frac{ln[1-\Pi(t)]}{1+s(t)} - \Pi(t). \tag{A2}$$

With weak selection ($s(t) \ll 1$), $\Pi(t)$ is of the same order as $s(t)$ and the above simplifies to

$$\frac{d\Pi}{dt} = -\frac{1}{2}\Pi(t)^2 + s(t)\ \Pi(t) + O(s^2) \tag{A3}$$

(Barton, 1995). This differential equation can be solved when selection on the recombinant haplotype varies according to $s(t) = s_a - p(t)(s_a - s_d)$ by first replacing the variable t with the variable p using the chain rule and $dp/dt = s_{net}p(1-p)$ (section 1 of Appendix S3). To leading order in the selection coefficients, the resulting solution for the fixation probability of the recombinant A_1B_0 haplotype is given by equation (2).

Appendix B

DERIVING THE DIFFUSION PROCESS WITH KILLING CONDITIONAL ON FIXATION OF THE A_1 ALLELE

Conditioning on the fixation of A_1 implies that either the A_1B_1 haplotype fixes (if the process is not killed) or the recombinant successfully establishes and leads to the fixation of the A_1B_0 haplotype (if the process is killed). Either way, the A_1B_1 haplotype cannot be lost while it is rare. We must thus adjust the drift term in the diffusion, $M(p)$, to account for the fact that the A_1B_1 haplotype will, on average, rise more rapidly when rare among those processes where the A_1B_1 haplotype is not lost. The variance term $V(p)$ and the killing term $K(p)$ are unchanged in the conditioned model, as these terms depend only on the current frequency of the A_1B_1 haplotype and not on its ultimate fate. From equation (9.5) in chapter 15 of Karlin and Taylor (1981), the conditional drift term $M^*(p)$ is given by.

$$M^*(p) = S_{net}\ p(1-p) + \frac{s(p)}{S(p)}\ p(1-p), \tag{B1}$$

where

$$s(p) = exp\left[-\int_0^p \frac{2M(\eta)}{V(\eta)} d\eta\right] \tag{B2}$$

$$S(p) = \int_0^p s(\xi)\, d\xi. \tag{B3}$$

Here, the values of $M(p)$ and $V(p)$ are for the unconditional diffusion process as outlined in the main part of the article. Plugging these terms into equations (B2) and (B3) and evaluating the integrals, we obtain the conditional drift term:

$$M^*(p) = S_{net}\ p(1-p)\ \frac{1+e^{-2pS_{net}}}{1-e^{-2pS_{net}}}. \tag{B4}$$

This revised drift term is then placed in equation (8), along with the variance and killing terms, which remain unchanged. Dividing the result by $p(1-p)$ yields equation (11) in the main text.

The conditional diffusion process requires some care, however, with the boundary conditions. The probability that the process is not killed given that the A_1B_1 haplotype is fixed remains one, $P^*(1) = 1$, as before. Conditioning assumes, however, that the $p = 0$ boundary is never reached. Rather than assigning $P^*(0)$, we instead assume that $P^*(p)$ varies little over very small values of p, given that the process will ultimately reach $p = 1$ if it is not killed. Thus, we use $dP^*(0)/dp = 0$ as a second boundary condition.

Solving equation (10), we find that the probability that the process is never killed, regardless of whether A_0 or A_1 ultimately fixes is

$$\begin{aligned} P(p) = &\left(U^0_{-\omega}[-2(pS_{net}+S_d)]\ \left(L^{-1}_{\omega}[-2S_d]\right.\right. \\ &\left. - L^{-1}_{\omega}[-2(S_{net}+S_d)]\right) - L^{-1}_{\omega}[-2(pS_{net}+S_d)] \\ &\left.\times \left(U^0_{-\omega}[-2S_d] - U^0_{-\omega}[-2(S_{net}+S_d)]\right)\right) \\ &/ \left(U^0_{-\omega}[-2(S_{net}+S_d)]\ L^{-1}_{\omega}[-2S_d]\right. \\ &\left. - U^0_{-\omega}[-2S_d]\ L^{-1}_{\omega}[-2(S_{net}+S_d)]\right), \end{aligned} \tag{B5}$$

whereas the solution to equation (11), conditioned on the fixation of the beneficial A_1 allele, given by B6 (below). Here, $U^b_a[z] = U[a,b,z]$ is the Tricomi confluent hypergeometric function, $L^{\alpha}_n[x]$ the generalized Laguerre polynomial (Abramowitz and Stegun, 1970), and ω is the compound parameter given by equation (6) in the main text. Additional details regarding the derivation and solutions for these equations are provided in a *Mathematica* 6.0 file (Supporting information, section 2).

$$P^*(p) = \left(\frac{1-e^{-2S_{net}}}{1-e^{-2pS_{net}}}\right) \times \frac{U^0_{-\omega}[-2(pS_{net}+S_d)]\ L^{-1}_{\omega}[-2S_d] - U^0_{-\omega}[-2S_d]\ L^{-1}_{\omega}[-2(pS_{net}+S_d)]}{U^0_{-\omega}[-2(S_{net}+S_d)]\ L^{-1}_{\omega}[-2S_d] - U^0_{-\omega}[-2S_d]\ L^{-1}_{\omega}[-2(S_{net}+S_d)]} \tag{B6}$$

Supporting Information

The following supporting information is available for this article:

Appendix S1. Testing the effect of recombination on the fixation of a linked, neutral allele.
Appendix S2. Methods used for multilocus simulations.
Appendix S3. Derivations in mathematica (HartfieldOttoSM.nb file available for download).
Figure S1. The mean frequency of a linked neutral allele following a successful selective sweep, given as a function of the recombination rate Nr between different sites.
Figure S2. The mean number of deleterious alleles that fix genome-wide following the completion of a successful selection sweep, given as a function of the recombination rate NR (see Fig. 4 for data presented relative to $S_a = 0$). $S_a = 0$ (black dashed line), 20 (black solid line), 40 (dark gray), or 80 (light gray).
Figure S3. Fixation probability of the advantageous mutant in multilocus simulations, as a function of the recombination rate NR. $S_a = 0$ (black dashed line), 20 (black solid line), 40 (dark gray), or 80 (light gray).
Figure S4. Fixation probability of a recombinant carrying two beneficial mutations.

Supporting Information may be found in the online version of this article.

Please note: Wiley-Blackwell is not responsible for the content or functionality of any supporting information supplied by the authors. Any queries (other than missing material) should be directed to the corresponding author for the article.